

TELEPHONE: 212-278-0400

```

; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-728-323A-2
Alignment Scores:
Pred. No.: 0.0194 Length: 1162
Score: 102.50 Matches: 36
Percent Similarity: 36.08% Conservative: 21
Best Local Similarity: 22.78% Mismatches: 59
Query Match: 7.43% Indels: 42
DB: 4 Gaps: 4

US-09-658-824-808 (1-781) x US-08-728-323A-2 (1-1162)
Qy 261 AGGCCGAGGAGTGGAGGGGCTCAGCGAAGCTGGTGGTCTCTGGGGTATCC:AGTCC 320
Db 422 LyslysgluaspGluGluaspGlyGlyAsnLysThrLeuSerIleGlnSer 441
Qy 321 CAGAAGCACCTGAACCC-----CGACAGAG 347
Db 442 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 461
Qy 348 ATTCTGACTCCCGACAGCGGACAGGAGGCGCATGAGCGACACACAAACACA 407
Db 462 ProLeuGlnGluProGln-----GlnGln 469
Qy 408 GAACACACAGCCAGTCC-----CAGAGCCCGAGTAATCGAGAG 446
Db 470 GlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 489
Qy 447 CCCCCAAAAGAAAGAACAGCAGCTGAAAGTCGGGATCCACCTGGGCGAGCAGACAGAA 506
Db 490 ProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
Qy 507 GAACATCAGATACAGCTGAGATCCAGTCGCGGACATGGAAGGTGATCGAAGAGCTG 566
Db 510 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 529
Qy 567 CATCAGTCAACACCGGGGATAAATCTGGATTGGTTCGCGCGTCAAGGTCAAGATAAT 626
Db 530 GlnGln-GlnGluPro-----GlnGlnGlnGln 538
Qy 627 ACCTAAGAGACACACTGTAAATGCCAAGCAGGTGAAGACCAACACA 678
Db 538 uProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 555

RESULT 2
US-08-821-818-1
; Sequence 1, Application US/08821818
; Patent No. 6146877
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: IDENTIFICATION OF THE PROGRESSION
; TITLE OF INVENTION: ELEVATED GENE-3 AND USES THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; TELEFAX: 212-391-0525
; APPLICATION NUMBER: US/08/821,818
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Chan, Albert Wai-Kit
; REGISTRATION NUMBER: 36,479
; REFERENCE/DOCKET NUMBER: 51523
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 457 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-821-818-1
Alignment Scores:
Pred. No.: 0.0509 Length: 457
Score: 97.00 Matches: 48
Percent Similarity: 37.99% Conservative: 20
Best Local Similarity: 26.82% Mismatches: 67
Query Match: 7.03% Indels: 44
DB: 4 Gaps: 9

US-09-658-824-808 (1-781) x US-08-821-818-1 (1-457)
Qy 112 AGTCAGAGTCAAGAGNACACAGCAGGCTCCGAGGGTGTGTGTCAGTACAGAGT 171
Db 220 ThrAspAsnLysAlaGluProSerGlySerHisSerArgValTrpGluTyrHisThrArg 239
Qy 172 GAGAAGGCCCTCGAAGTCGTCGCCCTCTCATGCGTGCACGCCCATCGACTTCTTGT 231
Db 240 GluArgProLysGlnGluGlyGluThrLysProGluGlnHisArgAlaGlyGlnSerHis 259
Qy 232 CTCGTCACGCCCATTAAGTACAGGAGGAGGAGGCCAGG-----AGTGGAGG 279
Db 260 ProCysGlnAsnAlaGluAlaGluGluGlyGlyProGluThrSerValCysSerGlySer 279
Qy 280 GCTCAGCGAAGCTGGGGTCTGTGGGGTATCCAGTCCAGAGACACCTGGGAACCCC 339
Db 280 AlaPheLeuLys-----AlaTrpValTyr-Arg-----ProGlyGluAs 292
Qy 340 GACAGAAGATTCGAGTCCCTCCCA-----GACGGGACCAGAGAGAGGAGCGCATGAGCA 393
Db 292 pThrGluGluGlu-GluAspSerAspLeuAspSerAlaGluGluAspThrAla----- 309
Qy 394 CACACACAAACACAGAACACACACAGC----- 419
Db 310 --HisThrCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProG 329
Qy 420 -----CAGTCCCAGGAGCCAGTAATGGAGAG---CCCCAAAAGAAAGAACACGAGCTG 471
Db 329 lyGluAspThrGluGluGluAspGlyAspTrpAspSerAlaGluGluAspAlaSerG 349
Qy 472 AAAGT---CGGATCCT-----ACACCTGGGC 495
Db 349 InSerCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyG 369
Qy 496 AGCAGACAGAAGATCAGATACAGCTGAGATCCAGTGCAGCGCGACATGGAA 548
Db 369 luAspThrGluGluGluAspSerGluAsnValAlaProValAspSerGlu 386

RESULT 3
US-08-821-818-5
; Sequence 5, Application US/08821818
; Patent No. 6146877
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: IDENTIFICATION OF THE PROGRESSION
; TITLE OF INVENTION: ELEVATED GENE-3 AND USES THEREOF
```

```

Db 349 lnSerCystThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyG 3699
Oy 496 AGCAGACAGACAGAAGACAGGATACAGCTCAGATCCGACGCGCACATGAA 548
: : |||||: : |||: :
Db 369 luAspThrGlucGlucLuAspSerGluAsnValaProValAspSerGlu 386
: : |||||: : |||: :
RESULT 4
US-08-893-852A-3
; Sequence 3, Application US/08893852A
; Patent No. 6080538
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,852A
; FILING DATE: Filed Herewith

```

: PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER:  
 : FILING DATE:  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Billings, Lucy J.  
 : REGISTRATION NUMBER: 36,749  
 : REFERENCE/DOCKET NUMBER: PF-0341 US  
 : TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 657 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; IMMEDIATE SOURCE:
;   LIBRARY: GenBank
;   CLONE: 53041
; US-08-893-852A-3

Alignment Scores:
Pred. No.: 0.0963      Length: 657
Score: 95.00      Matches: 61
Percent Similarity: 35.77%      Conservative: 27
Best Local Similarity: 24.80%      Mismatches: 93
Query Match: 6.88%      Indels: 65
DB: 3      Gaps: 12

US-09-658-824-808 (1-781) x US-08-893-852A-3 (1-657)

QY 55 GCTACTGACACACGGCGGTAGTGTCACAGGCAGATCCAACTGGGAGTTGAAGTGTGACT 114
||||||| ||: ::||| |||
Db 212 AlaThrGluGluYsThrGluAsnLysAlaAspProSerAsnProSer----- 229
||||||| ||: ::||| |||

QY 115 GAGACTGAAGAGAACACAGCAGCGCTCCGAGGAGGTTGTTGGTCAGTCACTCAGACTGAG 174
||||| ::||| |||
Db 230 -----SerGlySerHisSerArgAlaTrpGluTyrTyrSerArgGlu 243
||||| ::||| |||

QY 175 AAGCGCCTCGAAGTCGTGCTCCCTCTCATCGCGGTGCCACGCCCATGCACCTTCTTGTCTC 234
||||| ::||| |||

```





US-08-244-603A-1  
: Sequence 1, Application US/08244603A  
: Patent No. 6200768  
: GENERAL INFORMATION:  
: APPLICANT: Mandelkow, Eva-Maria  
: APPLICANT: Mandelkow, Eckhard  
: APPLICANT: Lichtenberg-Kraag, Birgit  
: APPLICANT: Biernat, Jacek  
: APPLICANT: Drewes, Gerard  
: APPLICANT: Steiner, Barbara  
: TITLE OF INVENTION: No. 6200768el Tools For The Diagnosis And  
: TITLE OF INVENTION: Treatment Of Alzheimer's Disease  
: NUMBER OF SEQUENCES: 1  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
: ADDRESSEE: Borun  
: STREET: 233 South Waker Drive, 6300 Sears Tower  
: CITY: Chicago  
: STATE: Illinois  
: COUNTRY: United States of America  
: ZIP: 60606-6402  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Tape  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/244,603A  
: FILING DATE:  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Joseph A. Williams, Jr.  
: REGISTRATION NUMBER: 38,659  
: REFERENCE/DOCKET NUMBER: 28384/32778  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 312-474-6300  
: TELEFAX: 312-484-0448  
: TELEX: 25-3856  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 441 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-244-603A-1

Alignment Scores:  
Pred. No.: 0.171 Length: 441  
Score: 92.00 Matches: 59  
Percent Similarity: 39.90% Conservative: 24  
Best Local Similarity: 28.37% Mismatches: 80  
Query Match: 6.67% Indels: 45  
DB: 4 Gaps: 11

US-09-658-824-808 (1-781) x US-08-244-603A-1 (1-441)  
QY 76 GGTCCACAGCAGATCAACTGGAGTTGAAGTGTGAGTGAGAGGAAACACGCA 135  
Db 100 GlyThrAlaGluGluAlaGlyIleGlyAspThrProSerLeuGluAspGluAla 119  
QY 136 GGCTTCGGAGGGTTGTGTGCTAGTCACTCAGTACAGAGGCGCTCGAAGTCGTC 195  
Db 120 Gly-----HisValThrGlnAla-ArgMetValSerLysSerLysAs 133  
QY 196 CCTCTCATCGGTGCCACGCCCATGGACCTCTTCTGTCGTCACGGCCATACTAGGAG 255  
Db 133 p-----GlyThrGlySerAsp-----As 139  
QY 256 GAGAGGGCGGAGGAGTGGAGGGCTCAGCGCA----- 289  
Db 139 pLysLysAlaLysGlyAlaAspGlyLysThrLysIleAlaThrProArgGlyAlaAlaPr 159

US-09-658-824-808 (1-781) x US-09-069-023-22 (1-221)  
QY 293 TGGGGTGTGTGGG-----GGTATC 313  
Db 93 TrpGlnHisValGlyProGlyTyrArgAspArgSerTyrAspProCysProGlyHis 112  
QY 314 CGAGTCCCAAGAGCCTGGACCCCGACAGAGATTCTGCATCCCGACGGGACGAC 373  
Db 113 TrpThrProGluAlaProSerSerGlyThrThrCysProGlyLeuProArgAlaSerGlu 132  
QY 374 GAGAGGGAC-----GGCATGAGCGACACACACAAACACACAGACCCAGTCCCA 427  
Db 133 GluGluGlyGlyProGluAspSerGluAla-ValGlnProArgThrProGluGlu 152  
QY 428 G-----GAGCCCGAGTAATGGA-----GA 445

US-09-069-023-22  
: Sequence 22, Application US/09069023A  
: Patent No. 6348573  
: GENERAL INFORMATION:  
: APPLICANT: Nunez, Gabriel  
: APPLICANT: Inohara, Naohiro  
: APPLICANT: Koseki, Takeyoshi  
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS  
: TITLE OF INVENTION: SIGNALING PATHWAY INHIBITORS AND ACTIVATORS  
: FILE REFERENCE: UM-03333  
: CURRENT APPLICATION NUMBER: US/09/069,023A  
: NUMBER OF SEQ ID NOS: 38  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 22  
: LENGTH: 221  
: TYPE: PRT  
: ORGANISM: Mus musculus  
: US-09-069-023-22

Alignment Scores:  
Pred. No.: 0.186 Length: 221  
Score: 90.50 Matches: 36  
Percent Similarity: 37.32% Conservative: 17  
Best Local Similarity: 25.35% Mismatches: 47  
Query Match: 6.56% Indels: 42  
DB: 4 Gaps: 6

US-09-658-824-808 (1-781) x US-09-069-023-22 (1-221)  
QY 293 TGGGGTGTGTGGG-----GGTATC 313  
Db 93 TrpGlnHisValGlyProGlyTyrArgAspArgSerTyrAspProCysProGlyHis 112  
QY 314 CGAGTCCCAAGAGCCTGGACCCCGACAGAGATTCTGCATCCCGACGGGACGAC 373  
Db 113 TrpThrProGluAlaProSerSerGlyThrThrCysProGlyLeuProArgAlaSerGlu 132  
QY 374 GAGAGGGAC-----GGCATGAGCGACACACACAAACACACAGACCCAGTCCCA 427  
Db 133 GluGluGlyGlyProGluAspSerGluAla-ValGlnProArgThrProGluGlu 152  
QY 428 G-----GAGCCCGAGTAATGGA-----GA 445

US-08-244-603A-1  
: Sequence 1, Application US/08244603A  
: Patent No. 6200768  
: GENERAL INFORMATION:  
: APPLICANT: Mandelkow, Eva-Maria  
: APPLICANT: Mandelkow, Eckhard  
: APPLICANT: Lichtenberg-Kraag, Birgit  
: APPLICANT: Biernat, Jacek  
: APPLICANT: Drewes, Gerard  
: APPLICANT: Steiner, Barbara  
: TITLE OF INVENTION: No. 6200768el Tools For The Diagnosis And  
: TITLE OF INVENTION: Treatment Of Alzheimer's Disease  
: NUMBER OF SEQUENCES: 1  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
: ADDRESSEE: Borun  
: STREET: 233 South Waker Drive, 6300 Sears Tower  
: CITY: Chicago  
: STATE: Illinois  
: COUNTRY: United States of America  
: ZIP: 60606-6402  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Tape  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/244,603A  
: FILING DATE:  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Joseph A. Williams, Jr.  
: REGISTRATION NUMBER: 38,659  
: REFERENCE/DOCKET NUMBER: 28384/32778  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 312-474-6300  
: TELEFAX: 312-484-0448  
: TELEX: 25-3856  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 441 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-244-603A-1

Alignment Scores:  
Pred. No.: 0.171 Length: 441  
Score: 92.00 Matches: 59  
Percent Similarity: 39.90% Conservative: 24  
Best Local Similarity: 28.37% Mismatches: 80  
Query Match: 6.67% Indels: 45  
DB: 4 Gaps: 11

US-09-658-824-808 (1-781) x US-08-244-603A-1 (1-441)  
QY 76 GGTCCACAGCAGATCAACTGGAGTTGAAGTGTGAGTGAGAGGAGCAACGCA 135  
Db 100 GlyThrAlaGluGluAlaGlyIleGlyAspThrProSerLeuGluAspGluAla 119  
QY 136 GGCTTCGGAGGGTTGTGGTCACTGACTCAGAGTGAAGAGCCCTCGAAGTCGTC 195  
Db 120 Gly-----HisValThrGlnAla-ArgMetValSerLysSerLysAs 133  
QY 196 CCTCTCATCGGTGCCACGCCCATGGACCTCTTCTGTCGTCACGGCCATACTAGGAG 255  
Db 133 p-----GlyThrGlySerAsp-----As 139  
QY 256 GAGAGGGCCGAGGAGTGGAGGGCTCAGCGCA----- 289  
Db 139 pLysLysAlaLysGlyAlaAspGlyLysThrLysIleAlaThrProArgGlyAlaAlaPr 159

QY 290 -----AGCTGGGTGCTCTTGGGGGTATCCGAGTC-----CCAGAGCACC 333  
Db 159 oProGlyGlnLysGlnAlaAsnAlaThrArgIleProAlaLysThrProAlaApr 179  
QY 331 TCGAACCCCGACACAGAGATTCTGGACTCCCGACAGCGGACGAGAGGCGCATGAG 390  
Db 179 oLysThrPro---ProSerSerGlyGluProProLysSerGlyAspArgSerGlyTyrSe 198  
QY 391 CGACACACACAAACACA---GAACCCACACAGCCAGTCCCGAGGACCCAGTAATCGAGAGC 447  
Db 198 r-SerProGlySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrP 218  
QY 448 CCAAAAAGAAGAACCA-----GCAGCTGAAAGTCGGGATCTCTACACCTCGGCAGCAGA 501  
Db 218 roProThrArgGluProLysLysValAlaValAlaArgThrProProLysSerProSerS 238  
QY 502 CAGAACAGATCAG---GATACAGCTGAGATCCAGTCCCGGACATGGAAGTGATGTC 558  
Db 238 erAlaLysSerArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal--- 256  
QY 559 AAGAGCTGCATCAGTCAACACACCGGGGATAAATCTGCGATTGGCTTCGGCGTCAAGTG 618  
Db 257 LysSerLysIleGlySerThrGluAsnLeuLysHisGlnProGlyGly---LysVal 275  
QY 619 AAGATAATACCTAAGAG 636  
Db 276 GlnIleIleAsnLysLys 281

RESULT 7  
US-09-069-023-22  
: Sequence 22, Application US/09069023A  
: Patent No. 6348573  
: GENERAL INFORMATION:  
: APPLICANT: Nunez, Gabriel  
: APPLICANT: Inohara, Naohiro  
: APPLICANT: Koseki, Takeyoshi  
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS  
: TITLE OF INVENTION: SIGNALING PATHWAY INHIBITORS AND ACTIVATORS  
: FILE REFERENCE: UM-03333  
: CURRENT APPLICATION NUMBER: US/09/069,023A  
: NUMBER OF SEQ ID NOS: 38  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 22  
: LENGTH: 221  
: TYPE: PRT  
: ORGANISM: Mus musculus  
: US-09-069-023-22

Alignment Scores:  
Pred. No.: 0.186 Length: 221  
Score: 90.50 Matches: 36  
Percent Similarity: 37.32% Conservative: 17  
Best Local Similarity: 25.35% Mismatches: 47  
Query Match: 6.56% Indels: 42  
DB: 4 Gaps: 6

US-09-658-824-808 (1-781) x US-09-069-023-22 (1-221)  
QY 293 TGGGGTGTGTGGG-----GGTATC 313  
Db 93 TrpGlnHisValGlyProGlyTyrArgAspArgSerTyrAspProCysProGlyHis 112  
QY 314 CGAGTCCCAAGAGCAGCTGGACCCCGACAGAGATTCTGGACTCCCGACGCGGACGAG 373  
Db 113 TrpThrProGluAlaProSerSerGlyThrThrCysProGlyLeuProArgAlaSerGlu 132  
QY 374 GAGAGGGAC-----GGCATGAGCGACACACACAAACACAGACCCAGTCCCA 427  
Db 133 GluGluGlyGlyProGluAspSerGluAla-ValGlnProArgThrProGluGlu 152  
QY 428 G-----GAGCCCGAGTAATGGA-----GA 445

```

152 uProGluLeuGluAlaGluAlaThrLysGlyAspGluProAspLeuGluGlnGluMetGI 172
QY 446 GCCCAAAAAGAAAGAACACGACGAGCTGAAAGTCGGGATCTCTACAGCTGGCGACGACAGACA 505
Db 172 uProGluProGluProGluValGluProGluProGluProGluProGluProGluProGI 192
QY 506 AGAAGATCAGGATACAGCTCAGATCCAGTCAGGCGACATGGAAGGTGATCTGCAAGAGCT 565
Db 192 uProGluProGluProGluProGluProGluProGluProGluProGluProGluProGlu 211
QY 566 GCATCAGTCAACACACGGGGATAATCTGATTGGTTTCGGGTCCTCGCGCTCAAGGTGAAGATAA 625
Db 212 -----GlyAspGluSer-----GluGlyCysGluAs 220
QY 626 TACC 629
Db 220 nThr 221

RESULT 8
US-08-893-852A-1
; Sequence 1, Application US/08893852A
; Patent No. 6080558
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,852A
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0341 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 674 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TMLR3DT01
; CLONE: 508302
US-08-893-852A-1

```

Alignment Scores:			
Pred. No.:	0.374	Length:	674
Score:	89.50	Matches:	53
Percent Similarity:	31.03%	Conservative:	28
Best Local Similarity:	20.31%	Mismatches:	79
Query Match:	6.49%	Indels:	102
DB:	3	Gaps:	8

US-09-658-824-808 (1-781) x US-08-893-852A-1 (1-674)

Qy	186	AGTCGTGTCCTCTCATCGGGTGCACGCCCATGACACCTTCTGTCTCGTCACGGCCAT	245
Db	249	SerProArgSerGlySerAspProArgSerTrpGluTyrArgSerGlyGluAlaSer	268
Qy	246	AACTAGGAGCAAGGAGGGCCGAGGAGTGCAGGGGCTCAGGCGCAAGCTGGG	296
Db	269	GluGluysGluGluysAlaHisGluGluThrGlyLysGlyGluAlaAlaProGlyPro	288
Qy	296	-----	296
Db	289	GlnSerSerAlaProAlaGlnArgProGlnLeuLysSerTrpTrpCysGlnProSerAsp	308
Qy	297	-----GTGCTGTTGGGGTATCCGAGTC	319
Db	309	GluGluSerGluValLysAlaLeuGlyAlaAlaGluLysAspGlyGluAlaGluCys	328
Qy	320	-----CCAGAAGCA	337
Db	329	ProProCysIleProProProSerAlaPheLeuLysAlaTrpValTyrTrpProGlyGlu	348
Qy	338	CCGACACAGAATCTGCATCTCCAGACGGGACGAGAGGACGGCATGACGCACACA	397
Db	349	AspThrGluGluGluAspGluGluAspGluAspSerAspSerGlySerAspGlu	368
Qy	398	CACAAACACACACACACAGCCAGTCCACAGGAGCCAGCATATGGA	443
Db	369	GluGlu-GlyGluAlaGluAlaSerSerThrProAlaThrGlyValPheLeuLysSe	388
Qy	443	-----	443
Db	388	rTrpValrTrpGlnProGlyGluAspThrGluGluGluAspGluAspSerAspThrGlu	408
Qy	444	-----CAGCCCCAAAAGAAAGAACCCAGCAGCTGAAGTCGGGATCCTACA	488
Db	408	ySerAlaGluAspGluGluAlaGluThrSerAlaSerThrProProAlaSerAlaPh	428
Qy	489	-----CCTGGGCACGACAGAGAAGAT	512
Db	428	eLeuLysAlaTrpValTyrArgProGlyGluAspThrGluGluGluGluAspGluAspVa	448
Qy	513	-----CAGGATACAGCTGAGATCCCAAGTCGCGGCACATGGAAGGTGATCT	556
Db	448	laspSerGluAspLysGluAspAspSerGluAlaAlaLeuGlyGluAlaGluSerAspPr	468
Qy	557	GCAAGAGCTGCATCAGTCAACACCGGGGATAATCTCGGATTGGGTTCGGCGCTCAAGG	616
Db	468	oHisProSerHisProAspGlnSerAlaHisPheArgGlyTyrpGlyTyr---ArgProGlu	487
Qy	617	TGAAGATAATACCTAAAGAGGAACACTGTAAATGCCACAGGACAGGTGAAGACCA	675
Db	487	ylsGluThr-GluGluGluGluAlaAlaGlu-----AspTrpGlyGluAlaGluPro	504

```

RESULT 9
US-08-159-969-2
; Sequence 2, Application US/08159969
; Patent No. 5492812
; GENERAL INFORMATION:
; APPLICANT: Voornels, Paul H.
; TITLE OF INVENTION: Diagnostic Method for Alzheimer's
; TITLE OF INVENTION: Disease
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,969
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/738,778
; FILING DATE: 01-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S.Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 4697-040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-159-969-2

Alignment Scores:
Pred. No.: 0.325 Length: 351
Score: 89.00 Matches: 58
Percent Similarity: 40.10% Conservative: 25
Best Local Similarity: 28.02% Mismatches: 79
Query Match: 6.45% Indels: 45
DB: 1 Gaps: 11

US-09-658-824-808 (1-781) x US-08-159-969-2 (1-351)
QY 76 GGTCCACAGCAGATCCAACTGGGAGTTGAAGTGTGAGTGAGAGTAAGAGAACACGACA 135
||| :|||::||| ||| :|||::|||
Db 42 GlyLeuLysAlaGluGlnAlaIleGlyAspThrProSerLeuGluAspGluAlaAla 61
QY 136 GCCTTCGGAGGGTGTGTGCTCACTCAGACTGAGAGGCCCTCGAAGTCGTCGTC 195
||| :|||::||| ||| :|||::|||
Db 62 Gly-----HisValThrGlnAla-ArgMetValSerLysSerLysAs 75
QY 196 CCTCTACGGTCCAGGCCCATCTCTCTGCTCGTACAGGCCATACTAGGGAG 255
||| |||
Db 75 p-----GlyThrGlySerAsp-----As 81
QY 256 GAAGGAGGGCGAGGAGTGGAGGCTCAGCGCA----- 289
|||::: |||::: |||::: |||::: |||:::
Db 81 pLysLysAlaLysGlyAlaAspGlyLysThrLysIleAlaThrProArgGlyAlaApr 101
QY 290 -----AGCTGGGTGCTGTTGGGGTATCCGAGTC-----CCAGAAGCACCC 330
||| |||
Db 101 oProGlyGlnLysGlyGlnAlaAsnAlaThrArgIleProAlaLysThrProAlaPr 121
QY 331 TGGAAACCCGACAAAGATTCTGGACTCCCGACAGCGGCACCAGGAGGGCGCATGAG 390
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 oLysThrPro---ProSerSerGlyGluProProlLysSerGlyAspArgSerGlyTyrrSe 140
QY 391 CGACACACAAAACACA---GAACACACACACAGTCCCAGGAGCCCAAGTAATGGAGAGC 447
||| ::| ||| |||::|||::||| ||| ||| |||
Db 140 r-SerProGlySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrP 160
QY 448 CAAAAAAGAAGAACCA-----CGACTGAAGTGGGATCCCTACACTGGGCGACGAGA 501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 roProThrArgGluProLysLysValAlaValAlaArgThrProProlLysSerProSers 180
QY 502 CAGAAGAAGATCAG---GATACACTGAGATCCCGAGTCCGCGACATGAAGTCATCTGC 558
||| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 erAlaLysSerArgLeuGlnThrAlaProValPrometProAspLeu-LysAsnVal--- 198
QY 559 AAGAGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTGTTGCGGCGTCAAGGTG 618
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

Db 75 p-----GlyThrGlySerAsp-----As 81  
Qy 256 GAAGGAGGCGGAGAGTGGAGGGCTCAGCGCA----- 289  
Db 81 pLysLysAlaLysGlyLysAlaAspGlyLysThrLysLysLeuAlaThrProArgGlyAlaAlaPr 101  
Qy 290 -----AGCTGGGTGCTGTTGGGGTATCCGAGTC-----CCAGAAGCACC 330  
Db 101 oProGlyGlnLysGlyGlnAlaAsnAlaThrArgLysProAlaLysThrProAlaPr 121  
Qy 331 TGGACCCCGACAGAGATTCTGCATCCCGAGGGGACCGAGAGGAGCGCATGAG 390  
Db 121 oLysThrPro---ProSerSerGlyGluProProLysSerGlyAspArgSerGlyTyrSe 140  
Qy 391 CGACACACACAACACACA-----GAACACACACAGCCAGTCCAGGAGCCAGTAATGGAGGC 447  
Db 140 r-SerProGlySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrP 160  
Qy 448 CCAAAAGAGAGACCA-----CGAGCTGAAGTCGGGATCCCTACACCTGGGAGCAGA 501  
Db 160 roProThrArgGluProLysLysValAlaValAlaArgThrProProLysSerProSer\$ 180  
Qy 502 CAGAGAGAGATCAG---GATACAGCTGAGATCCAGTCGCGCATGGAAGGTGATCTGC 558  
Db 180 erAlaLysSerArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal--- 198  
Qy 559 AAGAGCTGCATCAGTCAAAACACCGGGGATAAATCTGGATTGGGTTCGGCGTCAAGGTG 618  
Db 199 LysSerLysLeGlySerThrGluAsnLeuLysHisGlnProGlyGlyLys---LysVal 217  
Qy 619 AAGATAATACCTAAA 633  
Db 218 GlnlleValTyrLys 222

RESULT 11  
US-08-244-951A-10  
; Sequence 10, Application US/08244951A  
; Patent No. 5843779  
; GENERAL INFORMATION:  
; APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES  
; TITLE OF INVENTION: DIRECTED AGAINST THE MICROTUBULE-ASSOCIATED  
; TITLE OF INVENTION: PROTEIN TAU, HYBRIDOMAS SECRETING THESE  
; TITLE OF INVENTION: ANTIBODIES, ANTIGEN RECOGNITION BY THESE  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIERMAN & MUSERLIAN  
; STREET: 600 THIRD AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10016  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/244, 951A  
; FILING DATE: 19-JAN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP93/03499  
; FILING DATE: 10-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP/92/403403.6  
; FILING DATE: 14-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CHARLES A. MUSERLIAN  
; REGISTRATION NUMBER: 19,683  
; REFERENCE/DOCKET NUMBER: 410.003A

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 661-8000  
; TELEFAX: (212) 661-8002  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 391  
; TYPE: Amino Acid  
; STRANDEDNESS: Unknown  
; TOPOLOGY: Unknown  
; FEATURE:  
; NAME/KEY: mTHFMPH-taul fusion protein  
; US-08-244-951A-10  
Alignment Scores:  
Pred. No.: 0.339 Length: 391  
Score: 89.00 Matches: 58  
Percent Similarity: 40.10% Conservative: 25  
Best Local Similarity: 28.02% Mismatches: 79  
Query Match: 6.45% Indels: 45  
DB: 2 Gaps: 11  
US-09-658-824-808 (1-781) x US-08-244-951A-10 (1-391)  
Qy 76 GGTCCACAGCAGATCCAACTGGGAGTGAAGTGTGAGTGAAGAGGAGCACCACCA 135  
Db 81 GlyLeuLysAlaGluGluAlaGlyIleGlyAspThrProSerLeuGluAspGluAlaAla 100  
Qy 136 GGCCTCCGGAGGGTGTGTGTCAGTGAAGAGGCGCTCGAAGTCTGCTGCTC 195  
Db 101 Gly-----HisValThrGlnAla-ArgMetValSerLysSerLysAs 114  
Qy 196 CCTCTCTATGGCGTCCACGCCCATGGAGCTTCTTGTCTCAGCGGCATTAAGGAG 255  
Db 114 p-----GlyThrGlySerAsp-----As 120  
Qy 256 GAAGGAGGCGGAGGAGTGGAGGGCTCAGCGCA----- 289  
Db 120 pLysLysAlaLysGlyLysAlaAspGlyLysThrLysLysLeuAlaThrProArgGlyAlaAlaPr 140  
Qy 290 -----AGCTGGGTGCTGTTGGGGTATCCGAGTC-----CCAGAAGCACC 330  
Db 140 oProGlyGlnLysGlyGlnAlaAsnAlaThrArgLysProAlaLysThrProAlaPr 160  
Qy 331 TGGAAACCCCGACAGAGATTCTGCATCCCGAGCGGACCGAGAGGAGCGCATCAG 390  
Db 160 oLysThrPro---ProSerSerGlyGluProProLysSerGlyAspArgSerGlyTyrSe 179  
Qy 391 CGACACACACAACACACA-----GAACACACACAGCCAGTCCCGAGGAGCCAGTAATGGAGGC 447  
Db 179 r-SerProGlySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrP 199  
Qy 448 CCAAAAGAGAGACCA-----GCAGCTGAAAGTCGGGATCCCTACACCTGGGAGCAGA 501  
Db 199 roProThrArgGluProLysLysValAlaValAlaArgThrProProLysSerProSer\$ 219  
Qy 502 CAGAAGAGATCAG---GATACAGCTGAGATCCCGAGTCGCGCATGGAAGGTGATCTGC 558  
Db 219 erAlaLysSerArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal--- 237  
Qy 559 AAGAGCTGCATCAGTCAAAACACCGGGGATAAATCTGGATTGGGTTCGGCGTCAAGGTG 618  
Db 238 LysSerLysLeGlySerThrGluAsnLeuLysHisGlnProGlyGlyLys---LysVal 256  
Qy 619 AAGATAATACCTAAA 633  
Db 257 GlnlleValTyrLys 261

RESULT 12  
US-08-389-011-23  
; Sequence 23, Application US/08389011  
; Patent No. 5861257  
; GENERAL INFORMATION:  
; APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;



## RESULT 14

```
QY 196 CCTCTCATCGGTCGCCACGCCCATGGACCTTCTTGTCTCTCAGCGCCATACTAGGGAG 255
Db 114 p-----GlyThrGlySerAsp-----As 120
QY 256 GAAGGAGGGCCGAGGAGTCCAGGGCTCAGCGCA-----289
Db 120 pLysLysAlaLysGlyAlaAspGlyLysThrLysIleAlaThrProArgGlyAlaAlaPr 140
QY 290 -----ACGTGGGGTCTCTTGGGGGTATCCGAGTC-----CCAGAAGCACC 330
Db 140 oProGlyGlnLysGlyGlnAlaAsnAlaThrArgIleProAlaLysThrProProAlaPr 160
QY 331 TGGNACCCCCACAGAGATTCTGGACTCCCGACGCGGACGAGGAGGACGCGCATGAG 390
Db 160 oLysThrPro---ProSerSerGlyGluProProLysSerGlyAspArgSerGlyTyrSe 179
QY 391 CGACACACACAAACACA---GAACACACACAGCCAGTCCCGAGGAGCCCGAATGGAGAGC 447
Db 179 r-SerProGlySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrP 199
QY 448 CCAAAAAGAAGACCA-----GCAGCTGAAAGTCCGGGATCCTACACCTGGCGACGAGA 501
Db 199 roProThrArgGluProLysLysValAlaValAlaValArgThrProProLysSerProSerS 219
QY 502 CAGAAGAAGATCAG---GATACAGCTGAGATCCCGCGCGACATGGAAGGTGATCTGC 558
Db 219 erAlaLysSerArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal--- 237
QY 559 AAGAGCTGCATCAGTCAAAACACCGGGGATTAATCTGGATTGGTTCCCGCGGTCAAGGTG 618
Db 238 LysSerLysIleGlySerThrGluAsnLeuLysHisGlnProGlyGlyGly---LysVal 256
QY 619 AAGATAATACCTAA 633
Db 257 GlnIleValYrLys 261

RESULT 15
US-08-447-591-2
; Sequence 2, Application US/08447591
; Patent No. 5591440
; GENERAL INFORMATION:
; APPLICANT: CARMAN, WILLIAM
; APPLICANT: DECKER, RICHARD H
; APPLICANT: WALLACE, LESLEY
; APPLICANT: MIMMS, LARRY T
; APPLICANT: SOLOMON, LARRY R
; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR DETECTION
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,591
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/059,031
; FILING DATE: 07-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5347.US.01
; TELECOMMUNICATION INFORMATION:
```

```
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-447-591-2

Alignment Scores:
Pred. No.: 0.348 Length: 228
Score: 88.00 Matches: 45
Percent Similarity: 35.79% Conservative: 23
Best Local Similarity: 23.68% Mismatches: 59
Query Match: 6.17% Indels: 63
DB: 1 Gaps: 9

US-09-658-824-808 (1-781) x US-08-447-591-2 (1-228)

QY 744 GATAGTTTAAGTCAATATCTAATATAAAACACGCTTGCCTTGTGTTACGCTTGT----- 691
Db 52 AsnSerGlnSerGlnIleSerSerHisSerProThrCysCysProProIleCysProGly 71
QY 690 -----CTTCATTTAAACCTTGCTGTTGCTCTTACCTGCTCTTCTGTCATTTTACAGTG 640
Db 72 TyrArgTrpMetCysLeuArgPheIleIlePheLeuCysIle-----LeuLeuLeu 89
QY 639 TTCTCTTTTAGGTATTATCTTACCTTGACCGCGGACCCAAATCCACAGATTATCCCGG 580
Db 90 CysLeuIlePheLeuValLeuLeuAspTyrGlnGlyMetLeuProValCysPro--- 108
QY 579 TGTTTGACTGATGACGCTTTGCAGATCACCTTCCATGTGCG-----538
Db 109 LeuIleProGlySerSerThrThrSerThrGlyProCysArgAsnThrThrCysThrThr 128
QY 537 ---GCACCTGGGATCTCAGCTGATCTCTCTCTGCTGCTGCCAGGTGTAGG--- 483
Db 129 ProAlaGlnGlyThrSerMetPheProSerCysCys-CysThrLysProThrAspArgAs 148
QY 482 -----ATCCCGACTTTCAGCTGCTGTTCTTC-----456
Db 148 nCysThrCysIleProIleProSerSerTrpAlaPheAlaLysPheLeuTrpAl 168
QY 455 -----TTTTTGGGGCT 445
Db 168 aSerValArgPheSerTrpLeuSerLeuLeuValProPheValGlnTrpPheValGlyLe 188
QY 444 CTCCATTACTGGGCTCCTGGGACTG-----GCTGTGTGGTTCGTGTTGTGTGTGTGTCG 391
Db 188 uSerProThrValTrpLeuSerValIleTrpMetMetTrpTyr-----202
QY 390 CTCATGCCGTCCCTCTCTGTCGCCGTCCGGGA-----GTCCAGAACTTCTGTCGGG 337
Db 203 -----TrpGlyProSerLeuTyrAsnIleLeuSerPr 213
QY 336 GTTCCAGGTGCTTCTGGGACTCGGATAC 309
Db 213 oPheMetProLeuLeuProIlePheTyr 222

RESULT 16
US-08-450-943-2
; Sequence 2, Application US/08450943
; Patent No. 5593825
; GENERAL INFORMATION:
; APPLICANT: CARMAN, WILLIAM
; APPLICANT: DECKER, RICHARD H
; APPLICANT: WALLIS, LESLEY
; APPLICANT: MIMMS, LARRY T
; APPLICANT: SOLOMON, LARRY R
; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR DETECTION
```

```

Db      188  userProThrValTrpLeuSerValIleTrpMetMetTrpTyr----- 203
Qy      390  CTCAATGCCGTCCTCTCTGGTCCCGCTCGGGGA-----GTCCAGAACTCTCTGTCCGG 337
Db      203  -----TrpGlyProSerLeuTyTAsnIleLeuSerPr 213
Qy      336  GTTCCAGGTGCTTCTGGCACTCGGATAC 309
Db      213  oPheMetProLeuLeuProIlePheTyr 222

RESULT 17
US-08-059-031-2
; Sequence 2, Application US/08059031
; Patent No. 5595739
; GENERAL INFORMATION:
; APPLICANT: CARMAN, WILLIAM
; APPLICANT: DECKER, RICHARD H
; APPLICANT: WALLACE, LESLEY
; APPLICANT: MIMMS, LARRY T
; APPLICANT: SOLOMON, LARRY R
; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

1 OPENING SYSTEM: PC-DOS/MS-DOS
2 SOFTWARE: Patentin Release #1.0, Version #1.25
3
4 CURRENT APPLICATION DATA:
5 APPLICATION NUMBER: US/08/059,031
6 FILING DATE:
7 CLASSIFICATION: 424
8
9 ATTORNEY/AGENT INFORMATION:
10 NAME: POEEMBSKI, PRISCILLA E.

```

REGISTRATION NUMBER: 55,207  
REFERENCE/DOCKET NUMBER: 5347.US.01

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-059-031-2

Alignment Scores:
Pred. No.:      0.348
Score:          88.00
Percent Similarity: 35.79%
Best Local Similarity: 23.68%
Query Match:     6.17%
DB:              1

US-09-658-824-808 (1-781) x US-08-059-031-2 (1-228)

Qy    744   GATAGTTTAAAGTCAAATATCATTAATAAAAACGACTGGCTGTGTTTCAGCTTG----- 691
       ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    52   AsnSerGlnSerGlnIleSerHisSerProThrCysCysProPheLeuCysProGly 71
       ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy    690   -----CUTCAATTAAACTGTGGTGCCTCTTCACCGTCGCTTCGGCATTTTACAGTG 640
       ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    72   TyrArgtrpMetCysLeuArgArgPheIlePheLeuCysile-----LeuLeuLeu 89
       ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

539   TTECCCTCCTTTACCTAATTAATGCTACCTTCACCCCGGCAACCACAATATCACATTTATATCCCCG 580

```



```

Db 90 CysLeuilePheLeuLeuValLeuLeuAspTyrGlnGlyMetLeuProValCysPro--- 108
    |||      :|||      |||||      :|||      :|||      :|||
QY 579 TGTGTGACTGATCAGCTCTGTCAGATCACCCTCCATGTCGC----- 538
    :|||      :|||      :|||      :|||      :|||      :|||
Db 109 LeuileProGlySerSerThrThrSerThrGlyProCysArgAsnThrCysThrThr 128
    :|||      :|||      :|||      :|||      :|||      :|||
QY 537 ---GCACTGGGATCTCAGCTGTATCTGTCGTATCTTCTGCTGCTGCCAGGTGTAG-- 483
    |||      |||      :|||      :|||      :|||      :|||
Db 129 ProAlaGlnGlyThrSerMetPheProSerCysCys-CysThrLysProThrAspArgAs 148
    :|||      :|||      :|||      :|||      :|||      :|||
QY 482 -----ATCCGACTTTCAGCTGCTGCTGCTCTC----- 456
    |||||      |||      |||      |||      |||      |||
Db 148 nCysThrCysileProileProSerSerTrpAlaPheAlaLysPheLeuTrpGluTrpAl 168
    :|||      :|||      :|||      :|||      :|||      :|||
QY 455 -----TTTTGGGGCT 445
    :|||      :|||      :|||      :|||      :|||      :|||
Db 168 aSerValArgPheSerTrpLeuSerLeuLeuValProPheValGlnTrpPheValGlyLe 188
    :|||      |||      |||      |||      |||      |||
QY 444 CTCATTACTGGGCTCCTGGGACTG-----GCTGTGTGTTCTGTGTTrpPheValGlyLe 188
    |||      |||      |||      :|||      :|||      :|||
Db 188 uSerProThrValTrpLeuSerValileTrpMetMetTrpTyr----- 202
    :|||      :|||      :|||      :|||      :|||      :|||
QY 390 CTCATGCCGCTCCTCTCCTGCTGCCCTGCGGA-----GTCCAGAACTTCTGTGCGG 337
    :|||      :|||      :|||      :|||      :|||      :|||
Db 203 -----TrpGlyProSerLeuTyrAsnileLeuSerPr 213
    :|||      :|||      :|||      :|||      :|||      :|||
QY 336 GTTCCAGGTGCTTCTGGGACTCGGATAC 309
    |||      |||||      :|||      |||
Db 213 oPheMetProLeuLeuProilePheTyr 222
    :|||      :|||      :|||      :|||      :|||      :|||
RESULT 18
US-08-450-942-2
; Sequence 2, Application US/08450942
; Patent No. 5925512
; GENERAL INFORMATION:
; APPLICANT: CARMAN, WILLIAM
; APPLICANT: DECKER, RICHARD H
; APPLICANT: WALLACE, LESLEY
; APPLICANT: MIMMS, LARRY T
; APPLICANT: SOLOMON, LARRY R
; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR DETECTION
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450.942
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5347.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

US-08-450-942-2

```

Alignment Scores:
Pred. No.: 0.348      Length: 228
Score: 88.00      Matches: 45
Percent Similarity: 35.79%      Conservative: 23
Best Local Similarity: 23.68%      Mismatches: 59
Query Match: 6.17%      Indels: 63
DB: 2      Gaps: 9

```

US-09-658-824-808 (1-781) x US-08-450-942-2 (1-228)

```

QY 744 GATAGTTTAACTAATATCTAATAAACCACCTGCTGTTTTCAGCTTGT----- 691
    :|||      :|||      :|||      :|||      :|||      :|||
Db 52 AsnSerGlnSerGlnileSerHisSerProThrCysCysProPheLeuCysProGly 71
    :|||      :|||      :|||      :|||      :|||      :|||
QY 690 -----CTTCATTAAACTTGTGTTGCTCTTACCTGCTTCTGCGCATTTTACAGTG 640
    :|||      :|||      :|||      :|||      :|||      :|||
Db 72 TyrArgTrpMetCysLeuArgArgPheIleIlePheLeuCysIle-----LeuLeuLeu 89
    :|||      :|||      :|||      :|||      :|||      :|||
QY 639 TTCCTCTTTAGGTATTATCTTACCTTGAGCGCGGAACCCAAATCCAGATTATATCCCGCG 580
    :|||      :|||      :|||      :|||      :|||      :|||
Db 90 CysLeuilePheLeuLeuValLeuLeuAspTyrGlnGlyMetLeuProValCysPro-- 108
    :|||      :|||      :|||      :|||      :|||      :|||
QY 579 TGTGTGACTGATGACGCTCTTGCAGATCACCCTTCCATGTCGC----- 538
    :|||      :|||      :|||      :|||      :|||      :|||
Db 109 LeuileProGlySerSerThrThrSerThrGlyProCysArgAsnThrCysThrThr 128
    :|||      :|||      :|||      :|||      :|||      :|||
QY 537 ---GCACTGGGATCTCAGCTGTATCTGCTGCTGCTGCTGCTGCCAGGTGTAG-- 483
    |||      |||      :|||      :|||      :|||      :|||
Db 129 ProAlaGlnGlyThrSerMetPheProSerCysCys-CysThrLysProThrAspArgAs 148
    :|||      :|||      :|||      :|||      :|||      :|||
QY 482 ----ATCCGACTTTCAGCTGCTGCTGCTCTC----- 456
    |||||      |||      |||      |||      |||      |||
Db 148 nCysThrCysileProileProSerSerTrpAlaPheAlaLysPheLeuTrpGluTrpAl 168
    :|||      :|||      :|||      :|||      :|||      :|||
QY 455 -----TTTTGGGGCT 445
    :|||      :|||      :|||      :|||      :|||      :|||
Db 168 aSerValArgPheSerTrpLeuSerLeuLeuValProPheValGlnTrpPheValGlyLe 188
    :|||      |||      |||      |||      |||      |||
QY 444 CTCATTACTGGGCTCCTGGGACTG-----GCTGTGTGTTCTGTGTTrpGTTGTGTCG 391
    |||      |||      |||      :|||      :|||      :|||
Db 188 uSerProThrValTrpLeuSerValileTrpMetMetTrpTyr----- 202
    :|||      :|||      :|||      :|||      :|||      :|||
QY 390 CTCATGCCGCTCCTCTCCTGCTGCCCTGCGGA-----GTCCAGAACTTCTGTGCGG 337
    :|||      :|||      :|||      :|||      :|||      :|||
Db 203 -----TrpGlyProSerLeuTyrAsnileLeuSerPr 213
    :|||      :|||      :|||      :|||      :|||      :|||
QY 336 GTTCCAGGTGCTTCTGGGACTCGGATAC 309
    |||      |||||      :|||      |||
Db 213 oPheMetProLeuLeuProilePheTyr 222
    :|||      :|||      :|||      :|||      :|||      :|||
RESULT 19
PCT-US94-05090-2
; Sequence 2, Application PC/TUS9405090
; GENERAL INFORMATION:
; APPLICANT: CARMAN, WILLIAM
; APPLICANT: DECKER, RICHARD H
; APPLICANT: WALLACE, LESLEY
; APPLICANT: MIMMS, LARRY T
; APPLICANT: SOLOMON, LARRY R
; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR DETEC
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-463-702A-2

Alignment Scores:
Pred. No.:      0.685      Length:      1208
Score:          88.00      Matches:     58
Percent Similarity: 34.67%      Conservative: 20
Best Local Similarity: 25.78%      Mismatches:  94
Query Match:      6.38%      Indels:     53
DB:               4      Gaps:       10

```

US-09-658-824-808 (1-781) x US-09-463-702A-2 (1-1208)	
QY	66 AC GCGGGGTAGGTCACAGCAGCATCCAACTGGGAGTTGAAGTGTCAGTGAGAGTCAAGA 125     :   :     :      :
Db	47 ThrLeuIysArgThrGlyGlnAlaGlyGlyLeu-----Arg 60     :   :     :      :
QY	126 GGACCAGCAGGCTCCGGAGGGTTGT--GTGGTCAGTGACTCAGAGTCAGAGGCCCT 182 :          :    :
Db	61 SerSerGluSerLeuProAlaAlaGluAlaProGluProArgCysTrpGlyPro 80 :          :    :
QY	183 CGAAGTCGTCCTCCCTCTCATCGGTGCACGCCCATGACCTTCCTGCTCGTCACGGC 242     :    :    :    :
Db	81 HisLeuAsnArgAlaAlaThrLysSerProGlnProThrProGlyArgSerArgGlnGly 100     :    :    :    :
QY	243 CAT-----AACTAGGGAAGCAGGGCCGAGGAGTGGAGGGCTCAGCGCAACTGGG 296 :    :    :    :    :    :
Db	101 SerValProAspTyrGlyGlnArgLeuLysAlaAsnLeuLysGlyThrLeuGlnAlaGly 120 :    :    :    :    :    :
QY	297 GTGCTGTGGGG-----GTATCCGAGTCCAGAACACCTGGGAACC 338       :    :    :    :
Db	121 ProAlaLeuGlyArgArgProTrpProLeuGlyArgAlaSerSerLysAlaSerThrPro 140       :    :    :    :
QY	339 CGACAGAAGATTCTGGACTCCCACGACGGACACGAGAGGGACGCATGACGCACACAC 398 :    :    :    :    :    :
Db	141 Lys-----ProProGlyThrGlyProValProSerPheAlaGluLys 154 :    :    :    :    :    :
QY	399 ACAACACAGAACCCACACAGCCAGTCCCAGGAGCCCCAGTAATGGAGAGCCCCAAAAGAA 458 :    :    :    :    :    :
Db	155 ValSerAspGluPro-----ProGlnLeuPro 163 :    :    :    :    :    :
QY	459 GAACACGACGCTGAAAGTCGGGATCCTACACCTGGSCAG-----CAGACAGAA 506       :    :    :    :    :
Db	164 GluPro-----GlnProArgProGlyArgLeuGlnHisLeuGlnAlaSer 178 :    :    :    :    :    :
QY	507 GAAGATCAGGATACAGCTGAGATCCCAGTCGCGCACATGGAAGAGGTGATCTGCAAGAGCTG 566 :    :    :    :    :    :
Db	179 LeuSerGlnArgLeuGlySerLeuAsp-----ProGlyTrpLeuGlnArgCys 194 :    :    :    :    :    :
QY	567 CATCAGTCAAACACCG-----GGGATAAATCTGGATTTCGGGTC 605 :    :    :    :    :    :

QY 567 CATCAGTCAAACACCG-----GGGATAAATCTGGATTGGGTTC 605  
||| |||::: ||| ||| ||| |||

```

Db 195 His-SerGluValProAspPheLeuGlyAlaProLysAlaCysArgProAspLeuGlySe 214
QY 606 CGCGCTCAAGTGAAGATAATACCTAAAGAGGAACACTGTAATAATCCCAAGCAGGTGA 665
Db 214 rGluGluSerGlnLeuIleProGlyGluSerAlaValLeuGlyProGlyAlaGlySe 234
QY 666 AGACCAACCAAA 678
Db 234 rGlnGlyProGlu 238
RESULT 21
5164481-2
Patent No. 5164481
; APPLICANT: LACROIX, MARTIAL; ZREIN, MAAN; DIONNE, GERVASIS
; TITLE OF INVENTION: PEPTIDES AND ANALOGUES AND MIXTURES
; FOR DETECTING AND ELICITING ANTIBODIES TO RUBELLA VIRUS
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/397,767
; FILING DATE: 23-AUG-1989
; SEQ ID NO:2:
; LENGTH: 277
5164481-2
Alignment Scores:
Pred. No.: 0.426 Length: 277
Score: 87.50 Matches: 40
Percent Similarity: 29.61% Conservative: 5
Best Local Similarity: 26.32% Mismatches: 55
Query Match: 6.14% Indels: 52
DB: 6 Gaps: 7
US-09-658-824-808 (1-781) x 5164481-2 (1-277)
QY 372 TGGTCCCGTCTGGGAGTCCA-----GAATCTTCTGTCGGGTTCCAGGT 328
Db 74 TrpSerArgAlaProProProGluGluArgGlnGluSerArgSerGlnThrProAla 93
QY 327 GCTTCTGGGACTCGATACCCCAACAGCACCC-----CAGCTTCGCCCTG 283
Db 94 ProLysProSerArgAlaProProGlnGlnProGlnProProArgMetGlnThrGlyArg 113
QY 282 AGCCCTCCACTCTCCGCGCCCTCTCCCTAGTATTGCGCGTGACGAGACAAGAAGG 223
Db 114 GlyGlySerAlaProArgProGluLeuGlyProThrAsnProPheGlnAlaVal 133
QY 222 TCCATGGCGTGGCAGCGATGAGAGGACGACACTTCGAGGCGCTTCTCAGTCTGAGT 163
Db 134 AlaArgGlyLeuArgProProLeuHisAspProAspThrGluAla-----148
QY 162 CACTGACCACACACCCCTCCGGAAGCCTGC-----TGGTTCTCTC-----124
Db 149 -----ProThrGluAlaCysValThrSerTrpLeuTrpSerGluGlyGln 163
QY 123 -----TTCACCTCACTCACTTCAACTCCC-----97
Db 164 GlyAlaValPheTyrArgValAspLeuHisPheThrAsnLeuGlyThrProProLeuAsp 183
QY 96 -----AGTTGGATCTGCCTGGACCTACCCCGCGTGTCTCAGTAGCGGAGAAGA 46
Db 184 GluAspGlyArgTrpAspProAlaLeuMetTyrAsnProCys-----197
QY 45 ATCCGACCTCAGGAGCCGCGAGTCGCGGCTCAC 12
Db 198 -----Gly-ProGluProProAlaHis 204
RESULT 22
US-08-127-499A-8
; Sequence 8, Application US/08127499A
; Patent No. 5510264
; GENERAL INFORMATION:
; APPLICANT: VAN ALSTYNE, Diane
; APPLICANT: SHARMA, Lawrence Rajendra

```

```

; TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED
; TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/127,499A
; FILING DATE: 28-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 51916/102/INBI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1063 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; US-08-127-499A-8
Alignment Scores:
Pred. No.: 0.735 Length: 1063
Score: 87.50 Matches: 40
Percent Similarity: 29.61% Conservative: 5
Best Local Similarity: 26.32% Mismatches: 55
Query Match: 6.14% Indels: 52
DB: 1 Gaps: 7
US-09-658-824-808 (1-781) x US-08-127-499A-8 (1-1063)
QY 372 TGGTCCCGTCTGGGAGTCCA-----GAATCTTCTGTCGGGTTCCAGGT 328
Db 74 TrpSerArgAlaProProProGluGluArgGlnGluSerArgSerGlnThrProAla 93
QY 327 GCTTCTGGGACTCGATACCCCAACAGCACCC-----CAGCTTCGCCCTG 283
Db 94 ProLysProSerArgAlaProProGlnGlnProGlnProProArgMetGlnThrGlyArg 113
QY 282 AGCCCTCCACTCTCCGCGCCCTCTCCCTAGTATTGCGCGTGACGAGACAAGAAGG 223
Db 114 GlyGlySerAlaProArgProGluLeuGlyProThrAsnProPheGlnAlaVal 133
QY 222 TCCATGGCGTGGCAGCGATGAGAGGACGACACTTCGAGGCGCTTCTCAGTCTGAGT 163
Db 134 AlaArgGlyLeuArgProProLeuHisAspProAspThrGluAla-----148
QY 162 CACTGACCACACACCCCTCCGGAAGCCTGC-----TGGTTCTCTC-----124
Db 149 -----ProThrGluAlaCysValThrSerTrpLeuTrpSerGluGlyGln 163
QY 123 -----TTCACCTCACTCACTTCAACTCCC-----97
Db 164 GlyAlaValPheTyrArgValAspLeuHisPheThrAsnLeuGlyThrProProLeuAsp 183
QY 96 -----AGTTGGATCTGCCTGGACCTACCCCGCGTGTCTCAGTAGCGGAGAAGA 46
Db 184 GluAspGlyArgTrpAspProAlaLeuMetTyrAsnProCys-----197
QY 45 ATCCGACCTCAGGAGCCGCGAGTCGCGGCTCAC 12

```

```

Db 198 -----Gly-ProGluProProAlaHis 204
|||||
RESULT 23
US-08-482-847-8
; Sequence 8, Application US/08482847
; Patent No. 5556757
; GENERAL INFORMATION:
; APPLICANT: VAN ALSTYNE, Diane
; APPLICANT: SHARMA, Lawrence Rajendra
; TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR
; TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR
; TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THERETO, AND USES THEREOF
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,847
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/127,499
; FILING DATE: 28-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 51916/104/INBI
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1063 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
US-08-482-847-8
Alignment Scores:
Pred. No.: 0.735 Length: 1063
Score: 87.50 Matches: 40
Percent Similarity: 29.61% Conservative: 5
Best Local Similarity: 26.32% Mismatches: 55
Query Match: 6.14% Indels: 52
DB: 1 Gaps: 7

US-09-658-824-808 (1-781) x US-08-482-847-8 (1-1063)
QY 372 TGGTCCCGCTCGGGAGTCCA-----GAATCTCTCTCGGGGTCCAGGT 328
|||||
Db 74 TrpSerArgAlaProProProGluGluArgGlnGluSerArgSerGlnThrProAla 93
QY 327 GCTTCTGGGACTGGATACCCCAACAGCACCC-----CAGCTTCGCCCTG 283
|||||
Db 94 ProLysProSerArgAlaProProGlnGlnProGlnProProArgMetGlnThrGlyArg 113
QY 282 AGCCCTCCACTCCTCGGCCCTCCTCTCCCTAGTTATGGCCGTGACGACAGAAAGG 223
|||||
Db 114 GlyLysSerAlaProArgProGluLeuGlyProThrAsnProPheGlnAlaVal 133
QY 222 TCCATGGCGGTGGCACCGCATGAGAGGACGACGACTTCGAGGCGCTTCTCACTCTGAGT 163
::: |||::: ||| ||| ||| |||

```

---

```

Db 134 AlaArgGlyLeuArgProProLeuHisAspProAspThrGluAla----- 148
QY 162 CACTGACCACACACCTCGGGAAGCCTGC-----TGGTTCCTC----- 124
||| |||||||
Db 149 -----ProThrGluAlaCysValThrSerTrpLeuTrpSerGluGlyGln 163
QY 123 -----TTCACCTCTCACTCACACTTCACTCCC----- 97
Db 164 GlyAlaValPheTyrArgValAspLeuHisPheThrAsnLeuGlyThrProProLeuAsp 183
QY 96 -----AGTTGGATCTGCCTGTGGACCTACCCGCGCTCTCTCAGTAGCGGAGAAAGA 46
||| ||| |||
Db 184 GluAspGlyArgTrpAspProAlaLeuMetTyrAsnProCys----- 197
QY 45 ATCCAGACCTCAGGACCCGAGTCGCGGCTCAC 12
||| |||||||
Db 198 -----Gly-ProGluProProAlaHis 204
|||||
RESULT 24
US-08-127-499A-1
; Sequence 1, Application US/08127499A
; Patent No. 5510264
; GENERAL INFORMATION:
; APPLICANT: VAN ALSTYNE, Diane
; APPLICANT: SHARMA, Lawrence Rajendra
; TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED
; TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/127,499A
; FILING DATE: 28-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 51916/102/INBI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 992 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
US-08-127-499A-1
Alignment Scores:
Pred. No.: 0.913 Length: 992
Score: 86.50 Matches: 40
Percent Similarity: 29.87% Conservative: 6
Best Local Similarity: 25.97% Mismatches: 52
Query Match: 6.07% Indels: 56
DB: 1 Gaps: 7

US-09-658-824-808 (1-781) x US-08-127-499A-1 (1-992)
QY 372 TGGTCCCGCTCGGGAGTCCA-----GAATCTCTCTCGGGGTCCAGGT 328
|||||
Db 73 TrpSerArgAlaProProProGluGluArgGlnGluSerArgSerGlnThrProAla 92
|||||

```



```
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UW 2637
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1148 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-882-046-4

Alignment Scores:
Pred. No.: 0.969 Length: 1148
Score: 86.50 Matches: 37
Percent Similarity: 35.03% Conservative: 18
Best Local Similarity: 23.57% Mismatches: 49
Query Match: 6.07% Indels: 53
DB: 4 Gaps: 10

US-09-658-824-808 (1-781) x US-08-882-046-4 (1-1148)
QY 585 CCCCCTGTTGACGTGACGTCTTCAGATCACCTTCACATGTCGCGACTGGGACT 526
Db 443 ProLeuCysGluValAspValAspLeuCysGluProSerProCysArgAsn---GlyAla 461
QY 525 CAGCTGTAT-----CCTGATCTTCTCTGTCTGCTGCC----- 492
Db 462 ArgCysTyrAsnLeuGluGlyAsp-TyrtCysAlaCysProAspAspPheGlyGly 481
QY 491 -AGGTGAGTACCGACTTTTCAGTCTGCTGCTCTCTCTTT-----TGGGCT 445
Db 481 sasncysSerValProArgGluProCysTrp-ArgGlyLeuInsSerAspArgTrpLeuA 501
QY 444 CTCATTACTGGCTCTCGGAC-----TCG-----CTGTGCTGTCTGTGT 403
Db 501 rgValArgArgGlyAlaTrpAspAlaTrpHisSerThrSerGlyValCysGlyProHisG 521
QY 402 TTGTGTGTGCTGCTCATGCC-----GTCCCTCTCTGCTGCCGCTGGGAGTCCA 352
Db 521 LyArgCysValSerGlnProGlyGlyAsnPheSerCysIleCysAspSerGlyPheThrG 541
QY 351 GAATCTTCTCGGGGTTC----- 332
Db 541 lyThrTyrCysHisGluAsnIleAspAspCysLeuGlyGlnProCysArgAsnGlyGlyT 561
QY 331 -----AGGTGCTTC-----TGGGACTCGGATA 310
Db 561 hcCysIleAspGluValAspAlaPheArgCysPheCysProSerClyTrpGluGlyGluL 581
QY 309 CCCCCAACAGACCCCA-----GCTTCGCTGAGCCCTCCACTCC 269
Db 581 euCysAspThrAsnProAsnAspCysLeuProAspProCysHisser 596

RESULT 27
US-08-072-610-2
; Sequence 2, Application US/08072610
; Patent No. 5532133
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
US-08-719-822B-2
; Sequence 2, Application US/08719822B
; Patent No. 5874527
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,822B
; FILING DATE: 09/30/96
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: pVMB3.3.1
US-08-719-822B-2
Alignment Scores:
Pred. No.: 1.18 Length: 1018
Score: 85.50 Matches: 37
Percent Similarity: 40.00% Conservative: 27
Best Local Similarity: 23.12% Mismatches: 57
Query Match: 6.20% Indels: 39
DB: 2 Gaps: 5

US-09-658-824-808 (1-781) x US-08-719-822B-2 (1-1018)
QY 311 ATCCGAGTCCGACGACCTGGACCCGACGAGAGATT----- 350
Db 551 ValGlnValProValAlaValGly-ProAlaGlnGluValProThrGluGluLeu-detG1 570
QY 351 -CTGAGCTCCCGACGCGGACGAGGACGCGGATGACCGACACACAAACACA-- 407
Db 570 nLeuGlnAspAspPheGluLeuGluGlyThrAlaGluAlaProGluGluGlyGluLe 590
QY 408 -----GAACACACACGACCGACGAGCCAGTAAATGAGAGAGCCCAAAA 454
Db 590 uValLeuGluGlyGluGlyGluProThrGluGluGluProArgGluGlyGluProThrG1 610
QY 455 AGAAGAACACGACGCTGAAATCGGGATCCTACACCT----- 491
Db 610 uGlyValProGluGluGluGluLeuGluAlaThrProGluAspAspPheGluLeuGluG1 630
QY 492 -----GGCAGCAGACAGAGAAGATCAGGA 517

US-09-092-458-2
; Sequence 2, Application US/09092458
; Patent No. 6231861
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,458
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/719,821
; FILING DATE: 09/30/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: pVMB3.3.1
US-09-092-458-2
Alignment Scores:
Pred. No.: 1.18 Length: 1018
Score: 85.50 Matches: 37
Percent Similarity: 40.00% Conservative: 27
Best Local Similarity: 23.12% Mismatches: 57
Query Match: 6.20% Indels: 39
DB: 4 Gaps: 5
```







Query Match: 6.09% Indels: 9  
DB: 3 Gaps: 3

US-09-658-824-808 (1-781) x US-08-545-860D-28 (1-1187)

QY 320 CCAGAAGCCTGGAAACCCGACAGAGATTCTGGACTCCCGACGGGACGAGGAGG 379  
||| ||||| :|: ||||| |||  
Db 502 ProAlaAlaProProGluGlyProArgSerThrGluProProArgHisProGluSer 521  
QY 380 GACGGCATGACGACACACACAAACACAGACAGCAGTCCCGAGGAGCCAGTAA 439  
||| ||||| :|: ||||| |||  
Db 522 LysGlySerSerAspSer-AlaThrSerGlnGluHisSerGluSerLysAspPro----- 539  
QY 440 TGGAGAGCCCCAAAA-----GAAGAACCCAGCAGCTGAAAGTCGGGATCCTACACC 490  
||| ||||| :|: ||||| |||  
Db 540 -----ProProLysSerSerSerLysAlaProArgAlaProProGluAlaProHisPr 557  
QY 491 TGGGCAGCAGACA---GAAGAAGATCAGGATACAGCTCCAGTCCCGACATGGA 547  
||| ||||| :|: ||||| |||  
Db 557 oGlyLysArgSerCysGlnLysSerProAlaGlnGluProProGlnArgGlnThrVa 577  
QY 548 AGGTGATCTGCAAGAGCTGCATCATCAGTCAAACACCGGGGATAAATCTGGATTGGTTCGG 607  
||| ||||| :|: ||||| |||  
Db 577 lGlyThrLysGlnProLysLysProValLysAlaSerAlaArgAlaGlySerArgThrSe 597  
QY 608 GCGTCAAGTGAA 620  
||| ||||| |||  
Db 597 rLeuGlnGlyGlu 601

RESULT 34

PCT-US94-04496-28

; Sequence 28, Application PC/TUS9404496

; GENERAL INFORMATION:

; APPLICANT: Croce, Carlo

; APPLICANT: Canani, Eli

; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods

; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias

; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &

; STREET: One Liberty Place, 46th floor

; CITY: Philadelphia

; STATE: Pennsylvania

; COUNTRY: USA

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; FILING DATE: 07-MAR-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/04496

; FILING DATE: 22-APR-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/10930

; FILING DATE: 09-DEC-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/327,392

; FILING DATE: 19-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/320,559

; FILING DATE: 11-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/062,443

; FILING DATE: 14-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/971,094

; FILING DATE: 30-OCT-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/888,839

; FILING DATE: 27-MAY-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/805,093

; FILING DATE: 11-DEC-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: DeLuca Esq., Mark

; REGISTRATION NUMBER: 33,229

; REFERENCE/DOCKET NUMBER: TJU-1262

; TELEPHONE: (215) 568-3100

; TELEFAX: (215) 568-3439

; INFORMATION FOR SEQ ID NO: 28:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1187 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-545-860D-28

Alignment Scores: 1.81 Length: 1187

Pred: No.: 84.00 Matches: 30

Percent Similarity: 42.86% Conservative: 15

Best Local Similarity: 28.57% Mismatches: 51

US-08-545-860D-28

; Sequence 28, Application US/08545860D

; Patent No. 6040140

; GENERAL INFORMATION:

; APPLICANT: Croce, Carlo

; APPLICANT: Canani, Eli

; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods

; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias

; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region

; NUMBER OF SEQUENCES: 94

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &

; STREET: One Liberty Place, 46th floor

; CITY: Philadelphia

; STATE: Pennsylvania

; COUNTRY: USA

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/545,860D

; FILING DATE: 07-MAR-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/04496

; FILING DATE: 22-APR-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/10930

; FILING DATE: 09-DEC-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/327,392

; FILING DATE: 19-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/320,559

; FILING DATE: 11-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/062,443

; FILING DATE: 14-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/971,094

; FILING DATE: 30-OCT-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/888,839

; FILING DATE: 27-MAY-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/805,093

; FILING DATE: 11-DEC-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: DeLuca Esq., Mark

; REGISTRATION NUMBER: 33,229

; REFERENCE/DOCKET NUMBER: TJU-1262

; TELEPHONE: (215) 568-3100

; TELEFAX: (215) 568-3439

; INFORMATION FOR SEQ ID NO: 28:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1187 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-545-860D-28

Alignment Scores: 1.81 Length: 1187

Pred: No.: 84.00 Matches: 30

Percent Similarity: 42.86% Conservative: 15

Best Local Similarity: 28.57% Mismatches: 51

```

Score: 84.00 Matches: 30
Percent Similarity: 42.86% Conservative: 15
Best Local Similarity: 28.57% Mismatches: 51
Query Match: 6.09% Indels: 9
DB: 5 Gaps: 3

US-09-658-824-808 (1-781) x PCT-US94-04496-28 (1-1187)
QY 320 CCAGAACACCTGGAACCCCGACAGAAAGATTCTGGACTCCCGACAGCGGACGAGAGG 379
    |||||
Db 502 ProAlaAlaProGluGluProArgSerThrGluProArgHisProGluSer 521
    |||||
QY 380 GAGGCGATGAGCAGACACACAAACACAGACACAGCCAGTCCCGAGGCCAGTAA 439
    |||||
Db 522 LysGlySerSerAspSer-AlaThrSerGlnGluHisSerGluSerLysAspPro----- 539
    |||||
QY 440 TGGAGAGCCCAAAAA-----GAAGAACACGAGCTGAAAGTCGGGATCCTACACC 490
    |||||
Db 540 -----ProProLysSerSerSerLysAlaProArgAlaProGluAlaProHisPr 557
    |||||
QY 491 TGGGCACGACACA---GAAGACATCAGGATACAGCTGAGATCCCGAGTCCGCGACATGGA 547
    |||||
Db 557 oGlyLysArgSerCysGlnLysSerProAlaGlnGlnGluProProGlnArgGlnThrVa 577
    |||||
QY 548 AGGTGATCTGCAGAGCTGCATCAGTCAAAACACCGGGGATAAATCTGGATTTCGGTTCCG 607
    |||||
Db 577 lGlyThrLysGlnProLysProValLysProValLysAlaSerAlaArgAlaGlySerArgThrSe 597
    |||||
QY 608 GCGTCAAGGTGAA 620
    |||||
Db 597 rLeuGlnGlyGlu 601

RESULT 35
US-09-658-824-808 (1-781) x US-08-320-559-26 (1-1210)
QY 320 CCAGAACACCTGGAACCCCGACAGAAAGATTCTGGACTCCCGACAGCGGACGAGAGG 379
    |||||
Db 525 ProAlaAlaProGluGluProArgSerThrGluProArgHisProGluSer 544
    |||||
QY 380 GAGGCGATGAGCAGACACACAAACACAGACACACAGCCAGTCCCGAGGCCAGTAA 439
    |||||
Db 545 LysGlySerSerAspSer-AlaThrSerGlnGluHisSerGluSerLysAspPro----- 562
    |||||
QY 440 TGGAGAGCCCAAAAA-----GAAGAACACGAGCTGAAAGTCGGGATCCTACACC 490
    |||||
Db 563 -----ProProLysSerSerSerLysAlaProArgAlaProGluAlaProHisPr 580
    |||||
QY 491 TGGGCACGACACA---GAAGACATCAGGATACAGCTGAGATCCCGACGCGACATGGA 547
    |||||
Db 580 oGlyLysArgSerCysGlnLysSerProAlaGlnGlnGluProProGlnArgGlnThrVa 600
    |||||
QY 548 AGGTGATCTGCAGAGCTGCATCAGTCAAAACACCGGGGATAAATCTGGATTTCGGTTCCG 607
    |||||
Db 600 lGlyThrLysGlnProLysProValLysProValLysAlaSerAlaArgAlaGlySerArgThrSe 620
    |||||
QY 608 GCGTCAAGGTGAA 620
    |||||
Db 620 rLeuGlnGlyGlu 624

RESULT 36
US-08-545-860D-26
; Sequence 26, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaan, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: No. 6040140ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION NUMBER: US/08/320,559
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,443
; FILING DATE: 14 MAY 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,094
; FILING DATE: 30-OCT-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,830
; FILING DATE: 27-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/805,093
; FILING DATE: 11-DEC-91
; ATTORNEY/AGENT INFORMATION:

```

APPLICATION NUMBER: US/08/545,860D  
FILING DATE: 07-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04496  
FILING DATE: 22-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/10930  
FILING DATE: 09-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/327,392  
FILING DATE: 19-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/320,559  
FILING DATE: 11-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/062,443  
FILING DATE: 14-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,094  
FILING DATE: 30-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/888,839  
FILING DATE: 27-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/805,093  
FILING DATE: 11-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca Esq., Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: TJU-1262  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1210 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-860D-26

Alignment Scores:  
Pred. No.: 1.83 Length: 1210  
Score: 84.00 Matches: 30  
Percent Similarity: 42.86% Conservative: 15  
Best Local Similarity: 28.57% Mismatches: 51  
Query Match: 6.09% Indels: 9  
DB: 3 Gaps: 3

US-09-658-824-808 (1-781) x US-08-545-860D-26 (1-1210)

QY 320 CCAGAAGCACCTGGAACCCGACAGAGATTCTGGACTCCCGACGAGGGACGAGAGAGG 379  
DB 525 ProAlaAlaProGluGluProArgSerThrGluProProArgHisProGluSer 544  
QY 380 GAGCGCATGACGACACACACAGAACACAGACCCAGTCCCGACGAGCCAGTAA 439  
DB 545 LysGlySerSerAspSer-AlaThrSerGlnGluHisSerGluSerLysAspPro----- 562  
QY 440 TGGAGAGCCCAAAAA-----CAAGAACACGAGCTGAAGTCGGGATCCTACACC 490  
DB 563 -----ProProLysSerSerSerLysAlaProArgAlaProProGluAlaProHisPr 580  
QY 491 TGGGCAGCAGACA---GAAGAAGATCAGGATACAGTACCGATCCCGACGACATGGA 547  
DB 580 oGlyLysArgSerCysGlnLysSerProAlaGlnGluProProGlnArgGlnThrVa 600  
QY 548 AGGTGATCTCGAAGACTCCATCAGTCAAAACACCGGGGATAATCTGGATTGGGTTCG 607  
DB 600 lGlyThrLysGlnProLysLysProValLysAlaSerAlaArgAlaGlySerArgThrSe 620  
QY 608 GCCTCAAGGTGAA 620

Db 620 rLeuGlnGlyGlu 624  
RESULT 37  
PCT-US94-04496-26  
Sequence 26, Application PC/TUS9404496  
GENERAL INFORMATION:  
APPLICANT: Croce, Carlo  
APPLICANT: Canaan, Eli  
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &  
ADDRESSEE: Norris  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04496  
FILING DATE:

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca Esq., Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: TJU-1242  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1210 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-04496-26

Alignment Scores:  
Pred. No.: 1.83 Length: 1210  
Score: 84.00 Matches: 30  
Percent Similarity: 42.86% Conservative: 15  
Best Local Similarity: 28.57% Mismatches: 51  
Query Match: 6.09% Indels: 9  
DB: 3 Gaps: 3

US-09-658-824-808 (1-781) x PCT-US94-04496-26 (1-1210)

QY 320 CCAGAAGCACCTGGAACCCGACAGAGATTCTGGACTCCCGACGAGGGACGAGAGAGG 379  
DB 525 ProAlaAlaProGluGluProArgSerThrGluProProArgHisProGluSer 544  
QY 380 GAGCGCATGACGACACACACAGAACACAGACCCAGTCCCGACGAGCCAGTAA 439  
DB 545 LysGlySerSerAspSer-AlaThrSerGlnGluHisSerGluSerLysAspPro----- 562  
QY 440 TGGAGAGCCCAAAAA-----GAAGAACACGAGCTGAAGTCGGGATCCTACACC 490  
DB 563 -----ProProLysSerSerSerLysAlaProArgAlaProProGluAlaProHisPr 580  
QY 491 TGGGCAGCAGACA---GAAGAAGATCAGGATACAGTACCGATCCCGACGACATGGA 547  
DB 580 oGlyLysArgSerCysGlnLysSerProAlaGlnGluProProGlnArgGlnThrVa 600  
QY 548 AGGTGATCTCGAAGACTCCATCAGTCAAAACACCGGGGATAATCTGGATTGGGTTCG 607  
DB 548 AGGTGATCTCGAAGACTCCATCAGTCAAAACACCGGGGATAATCTGGATTGGGTTCG 607

```
Db 600 lGlyThrLysGlnProLysLysProValLysAlaSerAlaArgAlaGlySerArgThrSe 620
QY 608 GCGTCAAGTGAA 620
Db 620 rLeuGlnGlyGlu 624
RESULT 38
US-08-447-591-3
; Sequence 3, Application US/08447591
; Patent No. 5591440
; GENERAL INFORMATION:
; APPLICANT: CARMAN, WILLIAM
; APPLICANT: DECKER, RICHARD H
; APPLICANT: WALLACE, LESLEY
; APPLICANT: MIMMS, LARRY T
; APPLICANT: SOLOMON, LARRY R
; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR DETECTION
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,591
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/059,031
; FILING DATE: 07-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5347.US.01
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-447-591-3
Alignment Scores:
Pred. No.: 1.18 Length: 228
Score: 44 Matches: 44
Percent Similarity: 35.26% Conservative: 23
Best Local Similarity: 23.16% Mismatches: 60
Query Match: 5.82% Indels: 63
DB: 1 Gaps: 9
US-09-658-824-808 (1-781) x US-08-447-591-3 (1-228)
QY 744 GATAGTTTAAGTCAATATCTAATATAAACGAGCTGGCTGTTTCACCTTGT----- 691
Db 52 AsnSerGlnThrGlnIleSerHisSerProThrCysCysProProIleCysProGly 71
QY 690 -----CTTCATTAAACTGTGGTTCCTCTCACCTGCTCTTGGCATTTCACAGTG 640
Db 72 TyrArgTipMetCysLeuArgArgPheIleIlePheLeuCysIle-----LeuLeuLeu 89
QY 639 TTCCTCTTTAGGTATTATCTTCACCTTGAGCGCGGAACCAATCCAGATTATCCCGG 580
Db 90 CysLeullePheLeuLeuValLeuLeuAspTyrGlnGlyMetLeuProValCysPro--- 108
QY 579 TGTTTGACTGATGACGCTCTTCGACATCACCTTCATGTGCG----- 538
Db 109 LeuIleProGlySerSerThrThrSerThrGlyProCysArgAsnThrThrCysThrThr 128
QY 537 ---GCACTGGGATCTCAGCTGTATCTGATCTTCTCTGCTGCTGCCAGGTAGG-- 483
Db 129 ProAlaGlnGlyThrSerMetPheProSerCysCys-CysThrLysProThrAspArgAs 148
QY 482 -----ATCCCGACCTTCAGCTGCTGGTCTTC----- 45
Db 148 nCysThrCysIleProIleProSerSerTrpAlaPheValLysPheLeuTrpGluTrpAl 168
QY 455 -----TTTTGGGGCT 445
Db 168 aserValArgPheSerTrpLeuSerPheLeuValProIleValGlnTrpPheAlaGlyLe 188
QY 444 CTCATTACTGGCTCTCTGGGACTG-----GCTGTGTGGTCTCTGCTTTGTGTGTGTCG 391
Db 188 userProThrValTrpLeuSerValIleTrpMetMetTrpTrp----- 202
QY 390 CTCATGCCGTCCCTCTCTCTGGTGGGGA-----GTCAGAATCTCTGTGCGG 337
Db 203 -----TrpGlyProSerLeuTrpAsnIleLeuSerPr 213
QY 336 GTTCCAGGTGCTTCTGGGACTCGGATAC 309
Db 213 oPheMetProLeuLeuProIlePheTyr 222
RESULT 39
US-08-450-943-3
; Sequence 3, Application US/08450943
; Patent No. 5593825
; GENERAL INFORMATION:
; APPLICANT: CARMAN, WILLIAM
; APPLICANT: DECKER, RICHARD H
; APPLICANT: WALLIS, LESLEY
; APPLICANT: MIMMS, LARRY T
; APPLICANT: SOLOMON, LARRY R
; TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,943
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5347.US.01
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

US-08-450-943-3

## Alignment Scores:

Pred. No.: 1.18 Length: 228  
Score: 83.00 Matches: 44  
Percent Similarity: 35.26% Conservative: 23  
Best Local Similarity: 23.16% Mismatches: 60  
Query Match: 5.82% Indels: 63  
DB: 1 Gaps: 9

US-09-658-824-808 (1-781) x US-08-450-943-3 (1-228)

Qy 744 GATAGTTTAACTCAATATCTAATAAAACACGCTTGGTGTTCAGCTTGT----- 691  
Db 52 AsnSerGlnThrGlnIleSerHisSerProThrCysCysProProIleCysProGly 71  
Qy 690 -----CTTCATTTAAACTTGTGTCTTTCACCTGCTTCTGGCATTACAGTG 640  
Db 72 TyrArgTrpMetCysLeuArgPheIleIleLeuCysIle-----LeuLeuLeu 89  
Qy 639 TTCCTCTTTAGGTATTATCTTCACTTGACCGCGGAAACCCAAATCCAGATTATCCCGG 580  
Db 90 CysLeuIlePheLeuLeuValLeuLeuAspTyrGlnGlyMetLeuProValCysPro--- 108  
Qy 579 TGTTCAGTGTAGTCTTTCAGATCATCCTTCCATGTCCG----- 538  
Db 109 LeuIleProGlySerSerThrThrSerThrGlyProCysArgAsnThrThrCysThrThr 128  
Qy 537 ---GCACGTGGATCTCAGCTGTATCTCTGCTGTCTGCTGCTGCCAGGTGTAGG-- 483  
Db 129 ProAlaGlnGlyThrSerMetPheProSerCysCys-CysThrLysProThrAspArgAs 148  
Qy 482 -----ATCCCGACTTTCAGTGTGCTTCTTC----- 456  
Db 148 nCysThrCysIleProIleProSerSerTrpAlaPheValIleValGlnTrpPheAlaGlyLe 188  
Qy 444 CTCCTACTAGTGGTCTCTGCGACTG-----GCTGTGTGGTCTGTGTGTGTGTGTGCG 391  
Db 188 userProThrValTrpLeuSerValIleTrpMetMetTrpTyr----- 202  
Qy 390 CTCATGCCGTCCCTCTCTCTGCTGCCGTGGGA-----GTCCAGATCTTCTGTGCGG 337  
Db 203 -----TrpGlyProSerLeuTyrAsnIleLeuSerPr 213  
Qy 336 GTTCCAGGTGCTTCTGGGACTCGGATAC 309  
Db 213 oPheMetProLeuLeuProIlePheTyr 222

## RESULT 40

US-08-059-031-3

Sequence 3, Application US/08059031

Patent No. 5595739

## GENERAL INFORMATION:

APPLICANT: CARMAN, WILLIAM

APPLICANT: DECKER, RICHARD H

APPLICANT: WALLACE, LESLEY

APPLICANT: MIMMS, LARRY T

APPLICANT: SOLOMON, LARRY R

TITLE OF INVENTION: HEPATITIS B VIRUS MUTANTS, REAGENTS AND METHODS FOR DETECTION

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: ABBOTT LABORATORIES D377/AP6D

STREET: ONE ABBOTT PARK ROAD

CITY: ABBOTT PARK

STATE: IL

COUNTRY: USA

ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/059,031  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: POREMBSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5347.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 228 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-059-031-3

## Alignment Scores:

Pred. No.: 1.18 Length: 228  
Score: 83.00 Matches: 44  
Percent Similarity: 35.26% Conservative: 23  
Best Local Similarity: 23.16% Mismatches: 60  
Query Match: 5.82% Indels: 63  
DB: 1 Gaps: 9

US-09-658-824-808 (1-781) x US-08-059-031-3 (1-228)

Qy 744 GATAGTTTAACTCAATATCTAATAAAACACGCTTGGTGTTCAGCTTGT----- 691  
Db 52 AsnSerGlnThrGlnIleSerHisSerProThrCysCysProProIleCysProGly 71  
Qy 690 -----CTTCATTTAAACTTGTGTCTTTCACCTGCTTCTGGCATTACAGTG 640  
Db 72 TyrArgTrpMetCysLeuArgPheIleIleLeuCysIle-----LeuLeuLeu 89  
Qy 639 TTCCTCTTTAGGTATTATCTTCACTTGACCGCGGAAACCCAAATCCAGATTATCCCGG 580  
Db 90 CysLeuIlePheLeuLeuValLeuLeuAspTyrGlnGlyMetLeuProValCysPro--- 108  
Qy 579 TGTTCAGTGTAGTCTTTCAGATCATCCTTCCATGTCCG----- 538  
Db 109 LeuIleProGlySerSerThrThrSerThrGlyProCysArgAsnThrThrCysThrThr 128  
Qy 537 ---GCACGTGGATCTCAGCTGTATCTCTGCTGTCTGCTGCTGCCAGGTGTAGG-- 483  
Db 129 ProAlaGlnGlyThrSerMetPheProSerCysCys-CysThrLysProThrAspArgAs 148  
Qy 482 -----ATCCCGACTTTCAGTGTGCTTCTTC----- 456  
Db 148 nCysThrCysIleProIleProSerSerTrpAlaPheValIleValGlnTrpPheAlaGlyLe 188  
Qy 444 CTCCTACTAGTGGTCTCTGCGACTG-----GCTGTGTGGTCTGTGTGTGTGTGTGCG 391  
Db 188 userProThrValTrpLeuSerValIleTrpMetMetTrpTyr----- 202  
Qy 390 CTCATGCCGTCCCTCTCTCTGCTGCCGTGGGA-----GTCCAGATCTTCTGTGCGG 337  
Db 203 -----TrpGlyProSerLeuTyrAsnIleLeuSerPr 213  
Qy 336 GTTCCAGGTGCTTCTGGGACTCGGATAC 309  
Db 213 oPheMetProLeuLeuProIlePheTyr 222

Search completed: October 18, 2002, 09:53:09  
Job time : 33.5 secs







178	Qy	GC	CCTCGAAGTCGTGCTCCTCCTCATCGCGTCCGACGCCCATGGACCTTCITGTCTCGTC	237
			:::	
345	Db	Pro	ProGluGlycInValProValMetArgInAspGluGlyGluAla	361
			:::	
238	Qy	AC	GGCCATAACTAGGCAGGAAGGAGGGCCGAGGAGTGGAGGGCTCAGCGCAAGTCGGG	297
		:::	:::	
362	Db	-----	SerSerGluGluGluGlyAspGluGluGlyGlySerGluGluGlu	376
			:::	
298	Qy	TG	CTGTGGGGGTATCCGAGTCCAGAA	339
			:::	
377	Db	-----	GluGlyAspProSer-GluGluAspSerGlyGluAspSerGlyAspGlyAla	394
			:::	
340	Qy	GAC	AGAGATTCCTGGACTCCCGACAGCGGACACGAGGAGCGGCTCAGCGACACACA	399
		:::	:::	
394	Db	r	SerGluGluAlaGlyAlaAlaSerGluGluAlaSer-GlyThrAlaGlyLeuGlyGlu	414
			:::	
400	Qy	CAA	CACAGAACCCACAC	429
			:::	
414	Db	lu	GluThrGlnProSerThrGluGlyLeuAspSerGlyProAlaGlySerGlnAlaGlnA	434
			:::	
430	Qy	AG	CCCGAGTAATGGAGAGCCGCCAAAAGAGAAAGAACACGACGCTAAAGTCGGGATCCTACAC	489
		:::	:::	
434	Db	sp	ThrGluAlaGluGluProGluGlyHisGlnGlyProGluSerProIleThrAlap	454
			:::	
490	Qy	CT	GGGCAGCAGACAGAGAAGATCAGGATACACCTCAGATCCCGAGTCGCGACATCGGAG	549
			:::	
454	Db	ro	GlnGluGluThrGluAspValSerGlu-----GluValProMetArgasp	469
			:::	
550	Qy	GT	GATCTCGAAGAGCTGCATCAGTCAACACCCGGGGATAACTGGATTGGGTCCCGC	609
			:::	
469	Db	-----	-----	469
610	Qy	GTC	AGGTGAAGATAATACCTTAAGAGGAACACTGTAAATGCCAGAAGCAGGTGAAGAG	669
			:::	
470	Db	-----	-----ArgSerHisIleGlu	475
			:::	
670	Qy	CA	ACCACAGATTAAATGAAGACAAG	695
			:::	
475	Db	ys	ThrLeuMetLeuAsnGluAspIys	483

## RESULT 2

A: Accession: A29573  
 A: Molecule type: mRNA  
 A: Residues: 1-230 <MIR>  
 A: Cross-references: GB:J02730; NID:g204476; PIDN:AAA41276.1; PID:g204477  
 A: Experimental source: submandibular gland  
 A: Superfamily: proline-rich protein  
 C: Keywords: saliva; submandibular gland

**Alignment Scores:**

Alignment scores:	
Pred. No.:	0.344
Score:	106.00
Percent Similarity:	40.29%
Best Local Similarity:	28.78%
Query Match:	2
DB:	2
Length:	230
Matches:	40
Conservative:	16
Mismatches:	50
Indels:	33
Gaps:	6

US-09-658-824-808 (1-781) x A29573 (1-230)

QY 306 GGGGTATCCAGTCCCAGAAAGCACCTCGAACC CGACAGAAGATTCTGGACTCCCCAGAC 365  
||| ||| |||::: ||||| ||| |||  
pb 58 GlySerGluGluGluGlnGlnGlnGlnGlnProThrGlnAlaGluAsnGlnGluPro 77

Qy	366	GGGACGAGGAGGCGGCATGAGCGACACACAAACACACACACCCAGTC	425
Db	78	AlaThr-----SerGlySerGluGluGlnGln 88	
Qy	426	CAGGAGCCCACT-----AATGGAGAGCCC-----	449
Db	89	GlnGluProThrGlnAlaGluAsnGlnGluProProAlaThrSerGlySerGluGlu 108	
Qy	450	---CAAAAGAACAGAACAGCA---GCTGAAAGTCGGGATCTTACCTCGGCAGCAGACA 503	
Db	109	GlnGlnGlnGlnGluProThrGlnAlaGluAsnGlnGluProProAlaThrSerGlySer 128	
Qy	504	GAAAGATCAGGATACAGCTGAGATCCCGTCGGCGACATGGAAGGTGATCTGCAAGAG 563	
Db	129	GluGluGlnGlnGlnGlnGlnGlu-----SerThrGlnAlaGluAsnGlnGlu 144	
Qy	564	CTGCATCAGTCAAAACACCGGGGATAAATCTGGATTTGGGTTCCGGCTCAAGGTGAAGAT 623	
Db	145	-----ProSerAspSerAlaGlyGluGlnGlnGluThrGlnProGluGlu 159	
Qy	624	AATACCTTAAGAGGAACACTCTAAANTGCCAAGACAGGTGAAGACCAACAA 678	
Db	160	GlyAsn-ValGluSerProProSerProGluAsnSerGlnGlnProGln 177	
RESULT 3			
A38647			
Glutamine/glutamate-rich protein form A, submandibular gland - rat			
C:Species: Rattus norvegicus (Norway rat)			
C:Date: 21-Feb-1992 #sequence_revision 10-Apr-1992 #text_change 24-Nov-1999			
C:Accession: A38647			
R:Cooper, L.F.; Elia, D.M.; Tabak, L.A.			
J. Biol. Chem. 266, 3532-3539, 1991			
A:Title: Secretagogue-coupled changes in the expression of glutamine/glutamic acid-ri-			
A:Reference number: A38647; MUID:91139639			
A:Accession: A38647			
A>Status: preliminary			
A:Molecule type: mRNA			
A:Residues: 1-246 <COO>			
A:Cross-references: GB:M58653; NID:g204481; PIDN:AAA1278.1; PID:g204482			
A>Note: the authors translated the codon GAT for residue 135 as Asn, and CGA for resi-			
C:Superfamily: proline-rich protein			
Alignment Scores:			
Pred. No.: 0.346 Length: 246			
Score: 106.00 Matches: 41			
Percent Similarity: 37.88% Conservative: 9			
Best Local Similarity: 31.06% Mismatches: 63			
Query Match: 7.68% Indels: 19			
DB: 2 Gaps: 4			
US-09-658-824-808 (1-781) x A38647 (1-246)			
Qy	306	GGGGTATCCGAGTCCCGAAGACACCTGGAAACCCCGACAGAGATTCTGGACTCCCGAC 365	
Db	74	GlySerGluGluGlnGlnGlnGlnGlnGluProThrGlnAlaGluAsnGlnGluProPro 93	
Qy	366	GGGACGAGGAGGAGCGGCATGCGGACACACAAACACAGACACCACAGCCAGTC 425	
Db	94	AlaThr-----SerGlySerGluGluGlnGlnGln 104	
Qy	426	CAGGAGCCCACTAATGGAGAGCCGCCAAAAGAAAGAACACAGCAGCTGGAAGTCGGGATCCT 485	
Db	105	GlnGluProThrGlnAlaGluAsnGln-----GluProAlaThrSerGlySerGlu 122	
Qy	486	ACACTTGGCGACAGACAGAA-----GAAGATCAGGATACAGCTGAGATCCCA 533	
Db	123	GluGlnGlnGlnGlnGlnGluProThrGlnAlaGluAspGlnGlnProProAlaThrSer 142	
Qy	534	GTCCCGGACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAACACACC-----GGG 584	
Db	143	GlySerGluGluGlnGlnGlnGlnGlnGluSerThrGlnAlaGluAsnGlnGluProSer 162	
Qy	585	GATAAATCTGGATTTGGGTTCCGGCTCAAGGTGAAGATAATACCTAAGAGGAACACTG 644	

Db 163 AspSerAlaGlyGluGlyGlnGlnThrGlnProGluGluGlyAsn-VaIGluSerProPr 182  
QY 645 TAAATGCCAGAGCGGTGAAGACGACCAACACAA 678  
Db 182 oSerSerProGluAsnSerGlnGluGlnProGln 193

RESULT 4  
B38647  
glutamine/glutamate-rich protein form B, submandibular gland - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Feb-1992 #sequence\_revision 10-Apr-1992 #text\_change 11-Jan-2000  
C:Accession: B38647

R:Cooper, L.F.; Elita, D.M.; Tabak, L.A.  
J. Biol. Chem. 266, 3532-3539, 1991  
A:Title: Secretagogue-coupled changes in the expression of glutamine/glutamic acid-rich  
A:Reference number: A38647; MUID:91139639  
A:Accession: B38647  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-255 <COO>  
A:Cross-references: GB:M58654; NID:g204483; PIDN:AAA1279.1; PID:g204484  
A:Note: the authors translated the codon GAT for residue 135 as Asn, and GAT for residue  
C:Superfamily: proline-rich protein

Alignment Scores:  
Pred. No.: 0.347 Length: 255  
Score: 106.00 Matches: 41  
Percent Similarity: 37.88% Conservatve: 9  
Best Local Similarity: 31.06% Mismatches: 63  
Query Match: 7.68% Indels: 19  
DB: 2 Gaps: 4

US-09-658-824-808 (1-781) x B38647 (1-255)

QY 306 GGGGTATCCGAGTCCAGAACACCTGGAAACCCGACAGAGATTCTGGACTCCCGAC 365  
Db 74 GlySerGluGluGlnGlnGlnGlnGlnProThrGlnAlaGluAsnGlnGluPro 93  
QY 366 GGGACAGGAGAGGCGGCATCAGCGACACACAAACACACAGAACCCACACCGAGTCC 425  
Db 94 AlaThr-----SerGlySerGluGluGlnGlnGln 104  
QY 426 CAGGAGCCAGTAATGGAGAGCCCAAGAGACACGACGCTGAAGTCCGGATCCT 485  
Db 105 GlnGluProThrGlnAlaGluAsnGln-----GluProAlaThrSerGlySerGlu 122  
QY 486 ACACCTGGCGAGCAGACAGAA-----GAAGATCAGGATACAGCTGAGATCCCA 533  
Db 123 GluGluGlnGlnGlnGlnGluProThrGlnAlaGluAspGlnGlnProAlaThrSer 142  
QY 534 GTGGCGACATGGAAGTGATCTGCAAGAGCTGCATCAGTCAACACACC-----GGG 584  
Db 143 GlySerGluGluGlnGlnGlnGlnGlnGlnSerThrGlnAlaGluAsnGlnGluProSer 162  
QY 585 GATAAATCTGGATTGGTTCCGGCGCTCAAGGTGAAGATAATACCTAAAGAGGAACACTG 644  
Db 163 AspSerAlaGlyGluGlyGlnGlnGlnGlnProGluGluGlyAsn-VaIGluSerProPr 182  
QY 645 TAAATGCCAGAGCGGTGAAGACCAACACAA 678  
Db 182 oSerSerProGluAsnSerGlnGluGlnProGln 193

RESULT 5  
A32469  
80K protein H precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 12-Oct-1989 #sequence\_revision 12-Oct-1989 #text\_change 21-Jul-2000  
C:Accession: A32469  
R:Sakai, K.; Hirai, M.; Minoshima, S.; Kudoh, J.; Fukuyama, R.; Shimizu, N.  
Genomics 5, 309-315, 1989  
A:Title: Isolation of cDNAs encoding a substrate for protein kinase C: nucleotide sequen  
A:Reference number: A32469; MUID:90007553

A:Accession: A32469  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-527 <SAK>  
A:Cross-references: GB:J03075; NID:g182854; PIDN:AAA52493.1; PID:g182855  
C:Genetics:  
A:Gene: GDB:PRKCSH; G19P1  
A:Cross-references: GDB:119961; OMIM:177060  
A:Map position: 19p13.2-19p13.2  
C:Keywords: phosphoprotein  
F:1-14/Domain: signal sequence #status predicted <SIG>  
F:524-527/Region: endoplasmic reticulum retention signal

Alignment Scores:  
Pred. No.: 0.365 Length: 527  
Score: 106.00 Matches: 52  
Percent Similarity: 37.74% Conservatve: 28  
Best Local Similarity: 24.53% Mismatches: 76  
Query Match: 7.68% Indels: 56  
DB: 2 Gaps: 8

US-09-658-824-808 (1-781) x A32469 (1-527)

QY 52 TCCGCTACTCAGACACGGCGGTAGTCCACAGCGACATCCAACTGGGAGTTGAAGTGTG 111  
Db 230 SerValThrGluLeu---GlnThrHisProGluLeuAspThrAspGlyAspGlyAlaLeu 248  
QY 112 AGTCAGAGTGAA-----GAGGAACACGACGAGGCTTC 141  
Db 249 SerGluAlaGluAlaGlnAlaLeuSerGlyAspThrGlnThrAspAlaThrSerPhe 268  
QY 142 CGGAGGTTGTGTGTCAGTGCAGTGCAGAGGAGGCCCTCGAAGTCGTCGCCCTCTC 201  
Db 269 TyrAspArgValTrpAlaAla-IleArgAspLysTyrArgSerGlu-----283  
QY 202 ATCGGTGCCACGCCCATGGACCTTCTGTCTCGTCACGCCCTAATACTAGGAGG---AA 258  
Db 284 -----AlaLeuProThrAspLeuProAlaProSerAlaProAspLeuThrGluProLy 301  
QY 259 GAGCGCCGAGAGTGGAGGGGCTCAGCGGAAGCTGGGTGCTGTTGGGGGTATCCGAGT 318  
Db 301 sGluGluGln-----304  
QY 319 CCAGAGACACCTGGAAACCCGACAGAGATTCTGGACTCCCGACGCGGACACGAGAG 378  
Db 305 -ProProValProSerSerProThrGluGluGlu-----GluGluGluGluGlu 321  
QY 379 GGACGGCATCAGCAGACACACAAACACACAGCCAGTCCCGAGGCCAGTA 438  
Db 321 uGluGluGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 341  
QY 439 ATGGAGAGCCCAAGAAAGAACACGACGAGCTGAAAGTCGGGATCCTACACCTGGGAGC 498  
Db 341 euSerProGlnProAlaSerProAlaGluGluAspLysMetProProTyAspGluG 361  
QY 499 AGCAGAGAGAGATCAGGATACA-----GCTGAGA 528  
Db 361 InThrGlnAlaPheIleAspAlaAlaGlnGluAlaArgAsnLysPheGluGluAlaGlu 381  
QY 529 TCCGAGTGCAGCATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAACACCGGGGATA 588  
Db 381 rgSerLeuLysAspMetGluGluSerIleArgAsnLeuGluGln-----G 396  
QY 589 AATCTGGATTGGTTCCGGCGCTCAAGGTGAA 620  
Db 396 luIleSerPheAspPheGlyProAsnGlyGlu 406

RESULT 6  
F75420  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: F75420

A:Map position: 1

US-09-658-824-808 (1-781) x F75420 (1-319)

## RESULT 7

A:Cross-references: GB:X00017; NID:g13737; PIDN:CAA24915.1; PID:g578828  
A:Note: the authors translated the initiation codon ATT for residue 1 as I  
C:Comment: The cytochrome b gene was isolated from a 20-kb maxicircle.  
C:Genetics:  
A:Genome: mitochondrion  
A:Genetic code: SGC6  
A:Start codon: ATT  
C:Function:  
A:Description: the net reaction catalyzed by the ubiquinol--cytochrome-c reductase with two hydrogen ions taken up from the mitochondrial matrix and four hydrogens taken up from the respiratory chain  
A:Pathway: oxidative phosphorylation  
C:Superfamily: cytochrome b; cytochrome b homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; membrane protein  
F:2-334/Domain: cytochrome b homology <CBH>  
F:2-202/Domain: cytochrome b6 homology <CB6>  
F:216-334/Domain: plastoquinol--plastocyanin reductase 17k protein homology







```

Db      170 LysGlnProGlyTyrTyrrProThrSerProThrGlnProGluGlnLeuGlnGlnProThr 189
          |||||   |||   |||||   |||
QY      525 GAGATCCCACTCGGCACATGGAGGTGATCTGAAGAGTGTCATCAGTCAAACACCGGG 584
          ::      |||       |||   |||||   |||
Db      190 GlnGlyGlnGlnArgGlnGlnProGlyGlnGlnLeuArgGlnGlnGlnGlnGly 209
          :|||   |||   |||||   |||
QY      585 GATAAATCTGGATTGGG 602
          ::|||   |||
Db      210 GlnGlnSerGlyGlnGly 215

RESULT 15
JN0689
glutenin, high-molecular-weight Ax2* chain precursor - wheat
C:Species: Triticum aestivum (common wheat)
C>Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 15-Nov-1999
C:Accession: JN0689
R:Anderson, O.D.; Greene, F.C.
Theor. Appl. Genet. 77, 689-700, 1989
A>Title: The characterization and comparative analysis of high-molecular-w
A:Reference number: JN0689
A:Accession: JN0689
A:Molecule type: DNA
A:Residues: 1-815 <AND>
A>Note: The authors translated the codon CTA for residue 11 as Val, CAT fo
C:Comment: The main wheat storage proteins are divided into two groups. Th
A families.
C:Genetics:
A:Gene: Ax2*
C:Superfamily: glutenin
F.1-2l/Domains: signal sequence #status predicted <SIG>
F.22-815/Product: glutenin, high-molecular-weight Ax2* chain #status predi
Alignment Scores:
Pred. NO.: 2.53 Length: 815
Score: 96.50 Matches: 45
Percent Similarity: 33.01% Conservative: 23
Best Local Similarity: 21.84% Mismatches: 73
Query Match: 6.99% Indels: 66
DB: 2 Gaps: 6

US-09-658-824-808 (1-781) x JN0689 (1-815)
QY      61 GAGACAGCGGGTAGTGTCACAGGCAGATCCAACTGGGAGTTGAAGTGTGAGTGAGAGT 120
          ::      |||   |||   |||||   |||
Db      50 GlnLeuArgAspValSerProGluLucysGlnProValGlyGlyGlyProValAlaArgGln 69
          :|||   |||
QY      121 GAAGAGAACAACAGCAGGCTCCCGAGGGTTGTGTGTCAGTGACTCAGAGTGAGAAGGCC 180
          |||||
Db      70 TyrGlutIn----- 72
QY      181 CTCGAAGTCGCTGCCTCTCATGCGGTGCCAGGCCCATGGACCTTCTGTCTGTCACG 240
          ::|||   |||||
Db      73 ---GlnValValPro----- 77
QY      241 GCCATAACTAGGAGGAAGAGGCGGCGAGGTGGAGG----- 279
          |||||   |||
Db      78 -----ProLysGlyGlySerPheTyrProGlyGluThr 89
QY      280 -----GCTCAGCGGAAGCTGGGGTCTGCTGTTGGGGGTATCCGAGT----- 318
          |||   ::|
Db      90 ProProGlnGlnLeuGlnGlnSerIleLeuTrpGlyIleProAlaLeuLeuArgArgTyr 109
          |||   |||   |||||
QY      319 CCCAGAAGCACCTGGACCCTCCACAGAGAGTCTGGACTCCC-----CAGACG 366
          |||   ::|||   |||
Db      110 TyrLeuSerValThrSerProGlnGlnValSerTyrProGlyGlnAlaSerSerGln 129
QY      367 GGACCAAGAGAGGGGACGGCATGACGCACACAC-ACAAACACAGAACACACACCCGATCC 425
          |||||   |||||   |||
Db      130 ArgProGlyGlnGlnGlnGluTyrTyrLeuThrSerProGlnGlnSerGlyGlnTrp 149
          |||||   |||||
QY      426 CAGGACCCCATGTAATGGAGAG-----CCCCAAAAGGAAGAACCA 464
          |||||   |||||

```

Db 150 GlnGlnProGlyGlnGlnSerGlyTyrProThrSerProGlnGlnSerGlyGln 169  
QY 465 GCAGCTGAAGTCCGGGATCTTACACCTGGGAGCAGACAGAGAAGATCAGGATACAGCT 524  
Db 170 LysGlnProGlyTyrProThrSerProTrpGlnProGlnLeuGlnGlnProThr 189  
QY 525 GAGATCCAGTGCAGCAGTGAAGGTGATCTGCAAGCTGCATCAGTCAAAACACCGGG 584  
Db 190 GlnGlyGlnGlnArgGlnGlnProGlyGlnGlnLeuArgGlnGlyGlnGlnGly 209  
QY 585 GATAAATCTGGATTGGG 602  
Db 210 GlnGlnSerGlyGlnGly 215  
RESULT 16  
JQ0431  
hypothetical 35.5K protein - Streptomyces fradiae transposon Tn4556  
C:Species: Streptomyces fradiae  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 21-Jul-2000  
C:Accession: JQ0431  
R:Stemieniak, D.R.; Slightom, J.L.; Chung, S.T.  
Gene 86, 1-9, 1990  
A:Title: Nucleotide sequence of Streptomyces fradiae transposable element Tn4556: a class 1 transposon  
A:Reference number: JQ0424; MUID:90185236  
A:Accession: JQ0431  
A:Molecule type: DNA  
A:Residues: 1-348 <SIE>  
A:Cross-references: GB:M29297; NID:g153509; PIDN:AA88561.1; PID:g1196912

Alignment Scores:  
Pred. No.: 2.63 Length: 348  
Score: 96.00 Matches: 40  
Percent Similarity: 34.51% Conservative: 9  
Best Local Similarity: 28.17% Mismatches: 52  
Query Match: 6.73% Indels: 41  
DB: 2 Gaps: 5

US-09-658-824-808 (1-781) x JQ0431 (1-348)

QY 348 TCTCTGTGGGTTCCAGGTGCTTCTGGGACATCGGATACCC----- 307  
Db 220 SerArgAlaGlyCysProThrAlaAlaGlySerLeuLeuProAlaProArgProAla 239  
QY 306 -----CCACAGCACCCCGAGCTT----- 289  
Db 240 SerSerAlaSerSerProGlnAlaAlaAlaProAlaAlaProSerAlaThrArgLeuPro 259  
QY 288 ---CGCTGAGCCCTCCACTCTCTGGCCCTCTCTCTCCCTAGT----- 247  
Db 260 ArgArgThrProSerAlaProArgProSerSerArgProGlyThrPro-Arg----- 279  
QY 246 TATGGCCGTGACGAGACAAAGGTCCATGGCGTGGCCGACCATGAGAGGACGACGAC 187  
Db 280 AlaAlaArgProProArgArgThrProGlyThrPro-Arg-----ProAlaAl 296  
QY 186 TTCGAGGGCTTCTCACTCTGAGTCACTGACACACACACACCCCTCCGAGACCCCTGCTGTC 127  
Db 296 aAlaArgAla-----ArgAlaProAlaGlyCys 305  
QY 126 CTCTTCATCTCACTCACTCACTCACTCCAGTTGGATCTGCTGTCGACCTACCGCGG 67  
Db 305 sSerProAlaArgArgThrProSerAlaProThrArgArgCysArgAlaAlaArgArg 325  
QY 66 TGCTCTAGTGGGAGAAAGATCCAGACCTCAGGACCCGAGTCCGCGGCTCAGAGTCT 7  
Db 325 gGlySerProArgProProAlaAlaArgProGlyArgGlnGlyThrArgArgAspSe 345  
QY 6 CGCC 3  
Db 345 rAla 346

RESULT 17  
T49822

hypothetical protein B24H17.80 [imported] - Neurospora crassa  
C:Species: Neurospora crassa  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49822  
R:Schulte, U.; Algn, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakati  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: 225022  
A:Accession: T49822  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-613 <SCH>  
A:Cross-references: EMBL:AL356815; GSPDB:GN00116; NCSP:B24H17.80  
A:Experimental source: BAC clone B24H17; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B24H17.80  
A:Map position: 6  
A:Introns: 19/2; 50/3; 581/3

Alignment Scores:  
Pred. No.: 3.03 Length: 613  
Score: 95.50 Matches: 42  
Percent Similarity: 38.65% Conservative: 21  
Best Local Similarity: 25.77% Mismatches: 62  
Query Match: 6.92% Indels: 38  
DB: 2 Gaps: 6

US-09-658-824-808 (1-781) x T49822 (1-613)

QY 260 GAGGGCCGAGGAGTGGAGGGGCTCAGCGCA---AGCTGGGTGCT----- 301  
Db 230 GLUGlyArgAspValGlyAsnThrAsnArgIleSerGlyAlaAspAspMet 249  
QY 302 -----GTTGGGGTATCCGAGTCCAGAGCACCTGGAAACCCCGACAGAGATTCTGGA 355  
Db 250 GluLeuMetGlyGlyAspAspMetAspGluMetGlyGlyArgAspLeuAspGluAsp 269  
QY 356 CTCCCAGACGCGGACAG-----GAGAGGCGGCGATGCGGCACACACAAACAC 406  
Db 270 IleProAspAlaAspGluGlyPheGlyTyrAspGlyAlaSerAsp-AspAspAsnG 289  
QY 407 AGAACACACAGCCAGTCCAGGAGCCAGTATGAGAGAGCCCAAAAGAGAACACGACG 466  
Db 289 uGluAspGluGlnGluGluAsnAspGlnAlaGluGluGluGluGluAlaTh 309  
QY 467 AGCTGAATCTCGGATCTACACCTGGCGGAGCAGACAGAGAGATCAGGATACAGCTGA 526  
Db 309 rThrGlu---ArgAsnThrThrThrGlyAspHisGlyAsnAspAspGluAspThrValGI 328  
QY 527 GATCCAGTGCAGCATGGAAGGTGATCTGCAAGAGGTGATCAGTCAAAACACCGGGGA 586  
Db 328 uAlaPro-----GlnAlaAlaSerThrAlaThrGlyTh 339  
QY 587 TAAATCTGATTTGGTTCCCGCGTCAAGGTGAAGATAATACCTAAAGAGGAACTGTA 646  
Db 339 rGlySerSerIleLeuHisProArgGlyArgAsnAspAsn----- 352  
QY 647 AAATGCCAAGCAGGTGAAGAGCAACCAAGTTTAATAGAACAGCTGAACACG 706  
Db 353 -----SerAsnIleHisGlnGlnGlnGlnGlnAr 363  
QY 707 CAAGCTG 713  
Db 363 gGluLeu 365

RESULT 18  
S10001

Myd116 protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 05-Nov-1999

C:Accession: S10001

R:Lord, K.A.; Hoffman-Liebermann, B.; Liebermann, D.A.

Nucleic Acids Res. 18, 2823, 1990

A:Title: Sequence of Myd116 cDNA: a novel myeloid differentiation primary response gene



A:Reference number: S10001; MUID:90251472

A:Accession: S10001

A:Molecule type: mRNA

A:Residues: 1-657 <LOR>

A:Cross-references: EMBL:X51829; NID:g53040; PIDN:CAA36128.1; PID:g53041

Alignment Scores:

Pred. No.: 3.36 Length: 657  
Score: 95.00 Matches: 61  
Percent Similarity: 35.77% Conservative: 27  
Best Local Similarity: 24.80% Mismatches: 93  
Query Match: 6.88% Indels: 65  
DB: 2 Gaps: 12

US-09-658-824-808 (1-781) x S10001 (1-657)

```
QY 55 GCTACTGAGACACGGCGGTAGGTCCACAGGCAGATCCAACTGGGAGTTGAAGTGTGAGT 114
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 AlaThrGluGluGlyGlyThrGluAsnLysAlaAspProSerAsnSerProSer----- 229

QY 115 GAGAGTGAAGGAACACAGCAGGCTCCGGAGGTGTGTGTCAGTCACTCAGAGTCAG 174
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 -----SerGlySerHisSerArgAlaTrpGluTyrTyrSerArgGlu 243

QY 175 AAGGCCCTCGAAGTCGTGCTCCCTCATGCGGTGCCAGCCCATGGACCTTCTGTGCTC 234
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 244 LysProLysGlnGluGlyGluAlaLysValGluAlaHisArgAlaGlyGlnGlyHisPro 263

QY 235 GTCAGCGCCATACTAGGAGGAGGAGGCGCGAGG-----AGTGGAGGG 279
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 264 CysArgAsnAlaGluAlaGluGlyGlyProGluThrThrPheValCysThrGlyAsn 283

QY 280 GCTCAGGGGAAGCTGGGTGTGTGGGGTATCCGAGTCCAGAACGACCTGGAACCCC 339
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 AlaPheLeuLys-----AlaTrpValTyr-Arg-----ProGlyGluAs 296

QY 340 GACAGAGATCTGACTCTC-----CCAGACGGGACCAGAGAGGAGCGGACGAGCGCA 393
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 296 pThrGluGluGlu-AspAsnSerAspSerAspSerAlaGluGluAspThrAla----- 313

QY 394 CACACAAACACACAAACACACAGCCAGTCC----- 425
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 --GlnThrGlyAlaThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProG 333

QY 426 -----CAGGAGCCCAATAATGGAGCCCAAAAGAGAACACAGCAGCTG 471
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 333 lyGluAspThrGluGluAspSerAspSerAlaGluGluAspThrAlaGlnT 353

QY 472 AAAGTCGGGATCCT-----ACACCTGGGCAGC 498
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 353 hrGlyAlaThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyGluA 373

QY 499 AGACAGAGAAGATCAGGATACAGCTGAGATCCAGTCGCGACATGGAAGGTGATCTGC 558
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 spThrGluGluGluAsnSer-----AspLeuAspSerAlaGluG 386

QY 559 AGAGCTGCATCAGTCAAC--ACCGGGGATAATCTGGATT-----GGGTCC 606
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 386 luAspThrAlaGlnThrGlyAlaThrProHisThrSerAlaPheLeuLysAlaTrpValT 406

QY 607 GCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAATGCCAAGACGAGTCAA 666
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 406 yrArgProGlyGluAsp--Thr-GluGluGluAsnSerAspLeuAspSerAlaGluGlu 424

QY 667 GAGCAACCAAA 678
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 425 AspThrAlaGln 428
```

RESULT 19

AF2074

ferrichrome-iron receptor [imported] - Anabaena sp. (strain PCC 7120)

C:Species: Anabaena sp.

A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002

A:Accession: AF2074

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AF2074

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-863 <R>

A:Cross-references: GB:BA000019; PIDN:BA073847.1; PID:g17131239; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: al12148

Alignment Scores:  
Pred. No.: 3.79 Length: 863  
Score: 94.50 Matches: 53  
Percent Similarity: 45.13% Conservative: 35  
Best Local Similarity: 27.18% Mismatches: 66  
Query Match: 6.85% Indels: 42  
DB: 2 Gaps: 11

US-09-658-824-808 (1-781) x AF2074 (1-863)

```
QY 115 GAGAGTGAAGAG-----GAACCAGCAGGCTCCGAGGGTTGTGTGTCAGTCACTCAG 168
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 AspAlaGluGluLeuValGluValThrGlyValArg-----IleAsnGln 72

QY 169 AGTGAGAAGGCCCTCGAAGTCGTGCTCCCTCTCATGCGGTGCCAGGCCATGGACCT--- 225
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 73 ThrGluLysGlyValGluValIleLeuGluThrAlaLysGlyAspAlaLeuLysProVal 92

QY 226 -----TCTGTGCTGTCACGCCCAT-----ACTAGGAGAGAA 258
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 93 GlnLysAsnGluGlyAsnLysLeuIleAspIleProAsnSerGlnLeuArgTyrGlu 112

QY 259 GGAGGGCC-----GAGGAGTGGAGGGCTCAGCGGAAGTGGGTGCTGTTGGGG 308
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 113 GlyGlyAspThrPheArgGlnGluLysProPheAlaGlyIleAlaGluValLeuVal--- 131

QY 309 GTATCCGAGTCCAGAACCCACTCGAA-----CCCCGA 341
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 132 ValAsnGlnAspAsnAsnThrIleGlnValThrValThrGlyGluThrGlyLeuProIle 151

QY 342 CAGAAGATTCTGGACTCCCAAGAGGGGACCAAGAGGAGGCGGATGAGCGACACACACA 401
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 ValGluLeuPheAspGlyAspGluGlyLeuIlePheGlyValThrProAlaThrThr 171

QY 402 AACACAGAACACACAGC---CAGTCCCAGGAGCCCAAGTAATGAGAGAGCCCCAAAGAA 458
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 AlaGlnGlnProGlnThrProGlnAlaGlnGluLysProAlaIleGluIleProGlnGlu 191

QY 459 GAACACGACGCTGAAGTCGGGATCCT-----ACACCTGGCGACACACACA--- 503
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 192 GluProAlaAlaGlnGlnAspGluProIleGluLeuValValThrGlyGlnGlnAsnGly 211

QY 504 -----GAAGAAGATCAGGATACAGCT-----GAGATCCCGATCCGCGACATCGAA 548
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 TyrArgValGlnAspAlaThrThrAlaThrLysThrAspThrProLeuArgAspIlePro 231

QY 549 GGTGATCTGCAAGAGCTGCATCAGTCAACACCCGGGATAAATCT 593
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 GlnSerIleGlnValValProArgGluValLeuGluAspArgAsn 246
```

RESULT 20

S55659

tegument protein 64 - equine herpesvirus 2

C:Species: equine herpesvirus 2

C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 08-Oct-1999

A:Accession: S55659

R:Telford, E.A.R.; Watson, M.S.; Alrd, H.C.; Perry, J.; Davison, A.J.

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

Pred. No.:	3.96	Length:	392
Score:	94.00	Matches:	47
Percent Similarity:	38.64%	Conservative:	21
Best Local Similarity:	26.70%	Mismatches:	79
Query Match:	6.81%	Indels:	29
DB:	2	Gaps:	5

US-09-658-824-808 (1-781) x T48360 (1-392)

QY 207 GTGCCACGCCCATGGACCTTCTTGTCCTCAGGCCATAACTAGGAGGAAGGAGGCC 266

101 PROPRIETARY - INFORMATION

Db 101 oSerGlyGlyAlaGluLysLys-----ProSerProA[aA] 114

827 CACCAGGACCCCGACAGAGATCTGGACTCCCGACGGACAGGAGAGGGACGGCA 386

QY 387 TGAGCGACACACACAACACAGAACCAACACAGCCAGTCCACGACCCCACTAATCCACAC

DD 134 uGIuLysLysGluGluGluGluLysLysAlaSerProProPro----- 150

```
Db      |||||:::|||||||:::|||||
15] -pro[ysgluserthrValValIenIvsthrIvalcwhfsgcrcl...clwcrclouy... 170
```

50/ GAAGATCAGGATA-----CAGCTGAGATCCCAGTGCGGAC 542

0v 543 ATCGAAGCTCATCTCCACACACCTCCCTC

db 190 sAspLeuValIleValLysGlyIleIleAspValLysGlnLeuThrProTyrLeuAsnGI 210

[illegible]

QY 652 -----CCAGAAGCAGGTGAAGAGCAACCAAGTT 681

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526

hypothetical protein K06A9.1a - *Caenorhabditis elegans*

C;Accession: T34433

**A;Description:** The sequence of *C. elegans* cosmid K06A9.

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Cross-references: EMBL:U80846; PIDN:AAC70889.1; GSPDB:GN00028; CESP:K06A9

A; Gene: CESP:K06A9.1a

A: Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2; 99

Alignment Scores: 4.24 Length: 1032  
 Pred. No.: 94.00 Matches: 72  
 Score: 33.70% Conservative: 19  
 Percent Similarity: 26.67% Mismatches: 101  
 Best Local Similarity: 6.59% Indels: 78  
 Query Match: 2 Gaps: 10  
 DB:

US-09-658-824-808 (1-781) x T34433 (1-1032)

QY 604 AACCCAAATCCAGATTATCCCGGTGTTGACTGATGCAGCTCTTGGAGATCACTTCC 545  
 Db 534 SerProAsnProSerSerSerProAlaSerThrGlySerThrIleThrSerGlySer 553  
 QY 544 ATGTGCG-----CGCACTGGGATCTCA 524  
 Db 554 SerSerIleValSerThrValSerGlySerThrValSerGlySerThrGlyThrSer 573  
 QY 523 GCTGTATCCTGATCTTCTCTGCTGCTGCCAGGTGTAGGATCCCGACTTTTCAGTCTCT 464  
 Db 574 GlnSerThrLeuAlaSerSerThrAlaThrProGlySerSerSerThrValProSerSer 593  
 QY 463 GGTTCCTCTTTTGGGCTCTCATTTACTGGCTCTGGGACTGGCTGTGGTCTCTGTG 404  
 Db 594 SerSerProGlnProSerSer-GlnSerProAlaProAsnThrGlySerThrThrProse 613  
 QY 403 TTTGTGTGTCTGCTCATCGCTCCCTCTCTCTG----- 370  
 Db 613 rGlnThrSerSerGlnSerProSerProSerMetAsnProSerSerSerThrProThrG 633  
 QY 369 -----TCCCGTCTGGGAGTCCAGATCTTCTCT 341  
 Db 633 ySerSerGlnSerThrIleThrProGluGlySerThrAlaSerProThrGlySerThr 653  
 QY 340 CGGG-----GTTCCAGGTCTCT 323  
 Db 653 rGlySerThrPheSerValAlaThrGluValThrSerGlnSerThrValProSerGlyse 673  
 QY 322 T-----GGGACTCGGATACCCCAACAGACCCAG-----CTTCGCTCAGCCCTCCAC 272  
 Db 673 rSerLeuGlyThrGlnSerThrAsnSerProSerProSerSerLeuSerProSerThr 693  
 QY 271 T-----CCTCGGCCCT 239  
 Db 693 rSerGlyMetSerThrLeuThrSerGluProSerProSerSerThrGlnSerSerGlyAl 713  
 QY 238 TGACGAGACAAGGTCCATGGCGTGGCCGATGAGAGGAGC----- 192  
 Db 713 aGlnSerThrLeuThrProSerProAsnProSerGln-SerThrSerSerLeuGlu 733  
 QY 191 --ACGACTTCGAGGCTCTCTCACTGAGTCACTACCCACACACCCCTCCGGAAGCCTG 134  
 Db 733 erSerThrSerGlyAlaThrThrSerSerGlySerAlaGlyThrMetThrSerPro 753  
 QY 133 CTGGTTCCTCTTCACTCTCACTCACTTCACTCCAGTTGGATCTGCTGTGGACCTA 74  
 Db 753 erGlnSerSer-----ValGlySerSerGlnGlySert 765  
 QY 73 CCGGC---CGTGTCTAGTACGGAGAGAAATCCAGACTCAGGA-----C 29  
 Db 765 hrSerProAlaAlaSerThrThrSerGlyGluMetThrSerGlnGlySerThrGlnThrP 785  
 QY 28 CCGAGTCGCGGCTCAGCTCCGCCC 3  
 Db 785 roGlySerSerValSerThrSerAla 793

RESULT 23  
 T34434  
 hypothetical protein K06A9.1a - Caenorhabditis elegans  
 C: Species: Caenorhabditis elegans  
 C: Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000

C: Accession: T34434  
 R: Geisel, C.; Gattung, S.  
 submitted to the EMBL Data Library, December 1996  
 A: Description: The sequence of C. elegans cosmid K06A9.  
 A: Reference number: 221525  
 A: Accession: T34434  
 A: Status: preliminary; translated from GB/EMBL/DBJ  
 A: Molecule type: DNA  
 A: Residues: 1-2232 <GEI>  
 A: Cross-references: EMBL:U80846; PIDN: AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a  
 A: Experimental source: strain Bristol N2; clone K06A9  
 C: Genetics:  
 A: Gene: CESP:K06A9.1a  
 A: Map position: X  
 A: Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2

Alignment Scores: 4.48 Length: 2232  
 Pred. No.: 94.00 Matches: 72  
 Score: 33.70% Conservative: 19  
 Percent Similarity: 26.67% Mismatches: 101  
 Best Local Similarity: 6.59% Indels: 78  
 Query Match: 2 Gaps: 10  
 DB:

US-09-658-824-808 (1-781) x T34434 (1-2232)

QY 604 AACCCAAATCCAGATTATCCCGGTGTTGACTGATGCAGCTCTTGGAGATCACTTCC 545  
 Db 534 SerProAsnProSerSerSerProAlaSerThrGlySerThrIleThrSerGlySer 553  
 QY 544 ATGTGCG-----CGCACTGGGATCTCA 524  
 Db 554 SerSerIleValSerThrValSerGlySerThrValSerGlySerThrGlyThrSer 573  
 QY 523 GCTGTATCCTGATCTTCTCTGCTGCTGCCAGGTGTAGGATCCCGACTTTTCAGTCTCT 464  
 Db 574 GlnSerThrLeuAlaSerSerThrAlaThrProGlySerSerSerThrValProSerSer 593  
 QY 463 GGTTCCTCTTTTGGGCTCTCATTTACTGGCTCTGGGACTGGCTGTGGTCTCTGTG 404  
 Db 594 SerSerProGlnProSerSer-GlnSerProAlaProAsnThrGlySerThrThrProse 613  
 QY 403 TTTGTGTGTCTGCTCATCGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 370  
 Db 613 rGlnThrSerSerGlnSerProSerProSerMetAsnProSerSerSerThrProThrG 633  
 QY 369 -----TCCCGTCTGGGAGTCCAGATCTTCTCT 341  
 Db 633 ySerSerGlnSerThrIleThrProGluGlySerThrAlaSerSerProThrGlySerThr 653  
 QY 340 CGGG-----GTTCCAGGTCTCT 323  
 Db 653 rGlySerThrPheSerValAlaThrGluValThrSerGlnSerThrValProSerGlyse 673  
 QY 322 T-----GGGACTCGGATACCCCAACAGACCCAG-----CTTCGCTCAGCCCTCCAC 272  
 Db 673 rSerLeuGlyThrGlnSerThrAsnSerProSerProSerSerLeuSerProSerThr 693  
 QY 271 T-----CCTCGGCCCT 239  
 Db 693 rSerGlyMetSerThrLeuThrSerGluProSerProSerSerThrGlnSerSerGlyAl 713  
 QY 238 TGACGAGACAAGGTCCATGGCGTGGCCGATGAGAGGAGC----- 192  
 Db 713 aGlnSerThrLeuThrProSerProAsnProSerGln-SerThrSerSerLeuGlu 733  
 QY 191 --ACGACTTCGAGGCTCTCTCACTGAGTCACTACCCACACACCCCTCCGGAAGCCTG 134  
 Db 733 erSerThrSerGlyAlaThrThrSerSerGlySerAlaGlyThrMetThrSerPro 753  
 QY 133 CTGGTTCCTCTTCACTCTCACTCACTTCACTCCAGTTGGATCTGCTGTGGACCTA 74  
 Db 753 erGlnSerSer-----ValGlySerSerGlnGlySert 765

QY 73 CCGCG---CCTGTCCTAGTACGGAGAAAGAAATCCAGACCTCAGGGA-----C 29  
 Db 765 hrSerProAlaAlaSerThrThrSerGlyGluMetThrSerGlnGlySerThrGlnThrP 785  
 QY 28 CCGAGTCGCGGGTTCACAGCTCCGCC 3  
 Db 785 roGlySerSerValSerThrSerAla 793

RESULT 24  
 S15720  
 glutenin high molecular weight chain 1A1 precursor - wheat  
 C:Species: Triticum aestivum (common wheat)  
 C:Date: 08-Jun-1994 #sequence\_revision 12-May-1995 #text\_change 20-Aug-1999  
 C:Accession: S15720  
 R:Halford, N.G.; Field, J.M.; Blair, H.; Urwin, P.; Moore, K.; Robert, L.; Thompson, R.;  
 submitted to the EMBL data library, July 1991  
 A:Description: Analysis of HMW glutenin subunit encoded by chromosome 1A of bread wheat  
 A:Reference number: S15720  
 A:Accession: S15720  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-830 <HAL>  
 A:Cross-references: EMBL:X61009; NID:g21742; PIDN:CAA43331.1; PID:g21743  
 C:Superfamily: glutenin

Alignment Scores:  
 Pred. No.: 4.62 Length: 830  
 Score: 93.50 Matches: 45  
 Percent Similarity: 30.41% Conservative: 21  
 Best Local Similarity: 20.74% Mismatches: 70  
 Query Match: 6.78% Indels: 82  
 DB: 2 Gaps: 6

US-09-658-824-808 (1-781) x S15720 (1-830)  
 QY 61 GAGACACGGCGGTAGGTCCACAGCGAGATCCAACTGGGAGTTGAAGTGTGAGTGAGAGT 120  
 Db 50 GlnLeuArgaspValSerProGluCysGlnProValGlyGlyGlyProValAlaArgGln 69  
 QY 121 GAAGAGGACACAGAGGCTTCGGAGGGTGTGTGGTCAGTCACTCAGATGAGAAGGCC 180  
 Db 70 TyrGluGln----- 72  
 QY 181 CTCGAAGTCGTCCTCTCATCGGTGCCACGCCCATGACCTTCTGTCTGCTCAGC 240  
 Db 73 ---GlnValValPro----- 77  
 QY 241 GCCATACTAGGAGGAGAGGAGGCGGAGAGTGGAGG----- 279  
 Db 78 -----ProLysGlyGlySerPheTyrProGlyGluThrThr 89  
 QY 280 -----GCTCAGCGCAAGCTGGGTGCTGTGGGG----- 309  
 Db 90 ProProGlnGlnLeuGlnGlnSerIleLeuTrpGlyIleProAlaLeuLeuArgArgTyr 109  
 QY 310 -----TATCCGAGTCCCGAGGAGCCAGTAAATGGAGAG----- 446  
 Db 110 TyrLeuSerValThrSerProGlnGlnValSerTyrTyrProGlyGlnAlaSerSerGln 129  
 QY 334 AACCCCGACAGAGATTCGACTCCCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 393  
 Db 130 ArgProGlyGln-----GlyGlnGlnProGlyGlnGlnGlnGlnGlnGlnGlnGln 144  
 QY 394 CACAC-ACAAACACAGAACACACAGCAGTCCCGAGGAGCCAGTAAATGGAGAG----- 446  
 Db 145 TyrLeuThrSerProGlnGlnSerGlyGlnTrpGlnGlnProGlyGlnGlnGlnGlnGln 164  
 QY 447 -----CCCCAAAGAGAACACACAGCAGTGAAGTCGGGATCCTACACCT 491  
 Db 165 TyrTyrProThrSerProGlnGlnSerGlyGlnGlnGlnProGlyTyrTyrProThrSer 184  
 QY 492 GGGCAGCAGACAGAGAGAGATCAGGATACAGTACAGTCCCGAGTCCCGAGTCCCGAGAGGT 551

Db 185 ProTrpGlnProGluGlnLeuGlnGlnProThrGlnGlnGlnArgGlnGlnProGly 204  
 QY 552 GATCTCAGAGAGTGCATCAGTCAACACACCGGGGATAAAATCTGATTTGGC 602  
 Db 205 GlnGlyGlnGlnLeuArgGlnGlnGlnGlnSerGlyGlnGly 221

RESULT 25  
 T29074  
 hypothetical protein SC1C2.25c - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
 C:Accession: T29074  
 R:Redenbach, M.; Kieser, H.M.; Denapaite, D.; Eichner, A.; Cullum, J.; Kinashi, H.;  
 Mol. Microbiol. 21, 77-96, 1996  
 A:Title: A set of ordered cosmids and a detailed genetic and physical map for the 8  
 A:Reference number: 220556; MUID:97000351  
 A:Accession: T29074  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1329 <RED>  
 A:Cross-references: EMBL:AL031124; NID:e1312893; PID:e1312918; PIDN:CAA19992.1  
 C:Genetics:  
 A:Note: SC1C2.25c

Alignment Scores:  
 Pred. No.: 4.78 Length: 1329  
 Score: 93.50 Matches: 46  
 Percent Similarity: 38.54% Conservative: 28  
 Best Local Similarity: 23.96% Mismatches: 80  
 Query Match: 6.78% Indels: 38  
 DB: 2 Gaps: 8

US-09-658-824-808 (1-781) x T29074 (1-1329)  
 QY 174 GAAGCCCTCGAAGTCGTCCTCTCATCGGTGCCAGCCCATGACCTTCTGTCT 233  
 Db 882 AspAlaGluGlnGlyArgArgPro---GlnLeuProProArg----- 894  
 QY 234 CGTCACGCCATACTAGGAGGAGGAGGCGCCAGGAGTGGGGCTCAGGCGAAGCT 293  
 Db 895 -----GlyGlyProArgAlaGluLeuProGlyGlyAsnProGlnPro 908  
 QY 294 GGGGTGTGTGGGGTATCCGAGTCCAGAACACCTCGAAGCCCGACAGAAAGATT--- 350  
 Db 909 ArgValProSerTrpSerAspGluAsnAlaGlnProProValProArgAlaSerLeuAsp 928  
 QY 351 -----CTGGACTCCCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 398  
 Db 929 AlaProArgGlyHisAspGluProAspSerSerArgThrAspArgThrProArgLeuAsp 948  
 QY 399 ACAACACAGAACACACACAGCCAGTCCCGAGGAGCCAGT-----AATGGAGAGCC 449  
 Db 949 GluArgGlnGlyProGlySerThrAlaGluMetProAlaValProArgPheGlyGluPro 968  
 QY 450 CAAAGAAGAACACAGCAGCTGAAAGTCGGGATCCT-----ACACCTGGGCGACGAG 500  
 Db 969 GlnSerProAlaAlaThrAlaGluPheAlaArgProAspPheAspAlaProAlaProArg 988  
 QY 501 ACAGAAGAAGATCAGATACAGCTGAGATC-----CCAGTCGCGACATCGAAGGTGAT 554  
 Db 989 ArgAspGlnSerGlnAspThrGlyGlnTrpAlaGlnProGlyGlnAsn---GlnTyrAsp 1007  
 QY 555 CTGAAGAGCTGCATCATGATCAACACCGGGGATAAATCTGATTTGGGTCCTCCGCGCTAA 614  
 Db 1008 AlaArgAsnGluTyrGluAspGlnTyrGlyGlnGlnSerGlnTyrGlyGlnAspGlnTyr 1027  
 QY 615 GGTGAAGAT-----AATACCTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 644  
 Db 1028 GlyGlnAspGlnTyrAlaProGlyGlnTyrGlyGlnAlaGly-ProGlyGlnAspGlnTyr 1047  
 QY 645 TAAATGCCAGAAGCAGGTGAAGAGCAACCAACAA 678



A: Introns: 42/3; 108/3; 141/3; 210/3; 232/3; 501/1  
C: Superfamily: Caenorhabditis elegans hypothetical protein Y7588A.13

Alignment Scores:  
Pred. No.: 5.47 Length: 540  
Score: 92.50 Matches: 40  
Percent Similarity: 40.85% Conservative: 18  
Best Local Similarity: 28.17% Mismatches: 56  
Query Match: 6.70% Indels: 28  
DB: 2 Gaps: 8

US-09-658-824-808 (1-781) x T27400 (1-540)

QY 279 GGCTCAGCGAAGCTGGG---GTGCTGTTGGGGTATCCAGTCCCGAAGACACCTGGAA 335  
Db 310 GYAlaGlyGluAlaGlySerIleThrValThrAlaSerSerSerGluLys----- 326  
QY 336 CCCCAGAGAAGATTCTGGACTCCCGACAGCGGGACAGGAGCGGCGCATGAGCGACA 395  
Db 327 -----SerPro---AlaAspSerGlnGlySerAla---SerPro 337  
QY 396 CACACAACACAGACACACAGCCAGTCCCGAGGCGCCAGTAAT----- 440  
Db 338 AspGluSerThrGluGluProSerGlnAspHisThrProGlnAspLeuProMetAspGln 357  
QY 441 GGAGAGCCCCAAAAGAAGAACAGCAGCAGCTGAAAGTCGGGATCCTACACCTGGCGCAG 500  
Db 358 AlaProGlnAlaGluAspLeuGluArgSerAspAlaAspProGlySerArg 377  
QY 501 ACAGAGAAGATCAGGATCAGTGTAGTCCCGAGTCCCGGACATGAAGGTGATCTGCAA 560  
Db 378 AspGluSerGlnAsnAspVal-----ThrAlaSerLeuAspGlySerThrGlu 393  
QY 561 GAGTGTGATCAGTCAACACACC-----GGGGATAAATCTGGATTGGTTCGGCGGT 611  
Db 394 GluAlaSerGlnAspHisThrGluPheValAlaAspThrProGlyAlaGlyAspGluSer 413  
QY 612 CAGAGTGAA-----GATTAATCTAAAGAGGAACACTGTAAATGCCAGAGCAGGTGA 665  
Db 414 ThrAspAspAlaLeuAsn-LeuProMetAspThrAspGlnAlaProProHisAlaGlyG 433  
QY 666 AGAG 669  
Db 433 uGlu 434

RESULT 29  
A34615  
C: Species: Rattus norvegicus (Norway rat)  
C: Date: 29-Jun-1990 #sequence\_revision 09-Oct-1992 #text\_change 04-Feb-2000  
C: Accession: A34615  
R: Haydock, P.V.; Dale, B.A.  
DNA Cell Biol. 9, 251-261, 1990  
A: Title: Filaggrin, an intermediate filament-associated protein: structural and function  
A: Reference number: A34615; MUID: 90274870  
A: Accession: A34615  
A: Status: Preliminary  
A: Molecule type: mRNA  
A: Residues: 1-625 <HAY>  
A: Cross-references: GB:M21759; NID:q204143; PIDN:AAA41161.1; PID:q204144  
A: Note: The authors translated the codon GAA for residue 568 as Gln  
C: Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology  
C: Keywords: EF hand; epidermis

Alignment Scores:  
Pred. No.: 5.53 Length: 625  
Score: 92.50 Matches: 48  
Percent Similarity: 34.22% Conservative: 16  
Best Local Similarity: 25.67% Mismatches: 82  
Query Match: 6.70% Indels: 41  
DB: 2 Gaps: 9

US-09-658-824-808 (1-781) x A34615 (1-625)

QY 141 CCGAGGGTTGTGTGCTCACTGACTCAGAGCTGAGAGGCGCTCGAAGTCTCGTCCCTCT 200  
Db 9 ProGlnArgGluAlaAlaArgAspSerSer---GluGlnAlaGlnSerArgArg---Thr 26  
QY 201 CATGCGGTGCCACGCCCATGGACCTTCTGTCGTACCGCCCATCACTAGGAGGAGG 260  
Db 27 GluThrIleSerArg-----GlyArgSerGlyHisSerThrGly----- 39  
QY 261 AGGCGCCAGAGTGGAGG-----GGCTCAGCGGAGGCT 293  
Db 40 ArgAlaHisGluAspArgHisGluGlnAlaThrAspArgSerAlaGlySerGlySerArg 59  
QY 294 GGGTCTCTGTGGGTATCCAGTCCCGAAGCACCCTGGAAACCCCGACAGAGATTCG 353  
Db 60 GlyGlyGlnAlaGly---SerHisSerGluSerGluAlaSerGlyGlyGln----- 75  
QY 354 GACTCCCCAGACGGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 413  
Db 76 -----AlaGlyArgArgGlyThrAlaAlaThrArgHisThrSerArgProGlu 91  
QY 414 CACAGCCAGTCCCGAGGCGCCAGTAATGAGAGCCCCCAAAAGAACACACGAGCTGAA 473  
Db 92 GlnSerProAspThrAlaGlyArgThrGlySerSerArgGlyGlnGlnSerAlaGlnArg 111  
QY 474 AGTCGGGATCTACACCTGGCGAGCAGACAGAAAGATCAGGATACAGCTGAGATCCCA 533  
Db 112 HisGlyAspSerThrProGlySerThrArgThrGlySerArgGlyArgGlyGluSerPro 131  
QY 534 -----GTCCGCGACATGAAGGTGATCTGCAA 560  
Db 132 AlaGlyGlnGlnSerProAspArgAlaArgHisIleGluSerArgArgGlyArgThrArg 151  
QY 561 GAGTGTGATCAGTCAACACACCAGGATTAATCTGGATTTGGTTCGGCGGT----- 611  
Db 152 GluAlaSerAlaSerGlnSerSerAspSerGluGlyHisSerGlyAlaHisAlaGlyIle 171  
QY 612 ---CAAGGTGAAGATAATACC 629  
Db 172 GlyGlnGlyGlnThrSerThr 178

RESULT 30  
QRHUT1  
Microtubule-associated protein tau, long splice form - human  
N: Alternate names: microtubule-binding protein tau; neurofibrillary tangle protein tau  
N: Contains: microtubule-associated protein tau type II; microtubule-associated protein tau  
C: Species: Homo sapiens (man)  
C: Date: 30-Jun-1990 #sequence\_revision 03-May-1996 #text\_change 22-Jun-1999  
C: Accession: J50370; A30217; JN0009; S03796; S26665; S26666; S26662; S17302; A43444;  
R: Goedert, M.; Spillantini, M.G.; Jakes, R.; Rutherford, D.; Crowther, R.A.  
Neuron 3, 519-526, 1989  
A: Title: Multiple isoforms of human microtubule-associated protein tau: sequences ar  
A: Reference number: J50370; MUID: 90380393  
A: Accession: J50370  
A: Molecule type: mRNA  
A: Residues: 1-441 <GOE>  
A: Note: six isoforms are found: the clone httau40 sequence is shown. Residues 45-73, the clone httau24 sequence lacks inserts 1 and 2; the clone httau37 sequence lacks ir  
R: Goedert, M.; Wischik, C.M.; Crowther, R.A.; Walker, J.E.; Klug, A.  
Proc. Natl. Acad. Sci. U.S.A. 85, 4051-4055, 1988  
A: Title: Cloning and sequencing of the cDNA encoding a core protein of the paired he  
A: Reference number: A30217; MUID: 88234557  
A: Accession: A30217  
A: Molecule type: mRNA  
A: Residues: 1-44,103-274,306-441 <GO2>  
A: Cross-references: GB:J03778; NID:9338684; PIDN:AAA60615.1; PID:9338685  
R: Lee, G.; Neve, R.L.; Kosik, R.S.  
Neuron 2, 1615-1624, 1989  
A: Title: The microtubule binding domain of tau protein.  
A: Reference number: JN0009; MUID: 90180482  
A: Accession: JN0009  
A: Molecule type: mRNA  
A: Residues: 1-44,103-274,306-441 <LEE>

```

Alignment Scores:
  Pred. No.:      5.97      Length:      441
  Score:          92.00     Matches:      59
  Percent Similarity: 39.90%  Conservative: 24
  Best Local Similarity: 28.37% Mismatches:    80
  Query Match:    6.67%     Indels:      45
  DB:             1         Gaps:       11

US-09-658-824-808 (1-781) x QRRUT1 (1-441)

```

**Alignment Scores:**

US-09-658-824-808 (1-781) x QRHUT1 (1-441)

Pred. No.:	6.01	Length:	492
Score:	92.00	Matches:	56
Percent Similarity:	35.81%	Conservative:	21
Best Local Similarity:	26.03%	Mismatches:	84
Query Match:	6.45%	Indels:	54
DB:	2	Gaps:	9

US-09-658-824-808 (1-781) x S49147 (1-492)

Qy	637	CCCTCTTAGGTATTATCTTCACCTTCAGCCGCCGAACCCAAATCCACATTTATCCCGGGT	578
Db	277	ProLeu ---LeuLeuAspSerProThrSerArgThrProProPro-----ProSer	292
Qy	577	TTTGACTGATGCAGCTCTTGACATCACCCTTCATCGCGACTCGGATCTCAGCTGTA	518
Db	293	CysSerSerAlaSerSerCysSerSerAlaSerSerCysSerSerAlaSerAlaAla	312
Qy	517	TCCTGATCTTCTCT-----GTCCTGCTGCCAGGTGTAGGATCCCGACTTTTCAGCTGCT	464
Db	313	SerThrProSerGlyThrProThrCysCysAlaSerAlaAlaAlaAlaLeu-ArgLeuLe	332
Qy	463	GGTCTCTTTTGGGGCTCTCCATTACTGGCTCTCTGGGACTGGCTGTGTGGTCTGTG	404
Db	332	uTyrglyThrGlyGlyAlaGluAspLeuLeuAlaProGlyAlaProCysAlaAla----	350
Qy	403	TTTGTGTGTCTGCTCATGCGGCTCTCTCTGCTGCTCGCTGGGAGTCAGAAATCTTC	344
Db	351	----CysSerSerAlaSerCysAlaasnAsnAlaPheAlaPheGlyProGluLeuSerSe	369
Qy	343	TGTCGGGGTTCCAGGTGCTTCTGGGACTCGGATACCCCAACAGCACCCAGCTGCGCT	284
Db	369	rLeu-----lleThrProLeuAlaIleGlnThrHisAs	380
Qy	283	GAGCCCTCCACTCTCTCGGCCCTCTCTCCCTAGTTATGTCGCGTGACGACAGAAG	224
Db	380	nPheAlaAlaValAlaAlaAlaAlaTyrr-----TyrArgSerGlnGlnGlnGlnG	397
Qy	223	GTCCATGGGGCTGGCACCCGATCAGAGGGACGAGGACTCGAGGGCGCTTCTCACTCTG	164
Db	397	nGlnGlnGlyLeuAlaPro-----	403
Qy	163	TCACTGACACACAAACCCCTCGGAAGCCTGCTGGTTCCTCTTCACTCTCACTCACTTC	104
Db	404	-----ProAlaGlnProPro-AlaProProSerAlaThr-----	414
Qy	103	AACTCCCACTTGATCTGCCTGTGGACCTACC-----	CGCC 68
Db	415	--LeuProAlaGlyAlaAlaAlaProProSerProPheSerPheGlnLeuProArgA	434
Qy	67	GTGTCTCTAGTAGCGGAGAAAGATCCAGACCTCAGGACCC	27
Db	434	rdLeuSerAspSerProValPheAspAlaProProSerPro	447

RESULT 32  
B35962  
protein-tyrosine kinase (EC 2.7.1.112) I, splice form A -- human  
N:Alternate names: c-abl proto-oncogene homolog 2  
C:Species: Homo sapiens (man)  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 04-Feb-2000  
C:Accession: B35962  
R:Kruh, G.D.; Perego, R.; Miki, T.; Aaronson, S.A.  
Proc. Natl. Acad. Sci. U.S.A. 87, 5802-5806, 1990  
A:Title: The complete coding sequence of arg defines the Abelson subfamily of cytoplasmic

C; Keywords: alternative splicing; ATP; kinase-related transforming protein; phospho  
F; 78-126/Domain: SH3 homology <SH3>  
F; 137-227/Domain: SH2 homology <SH2>  
F; 250-510/Domain: protein kinase homology <KIN>  
F; 258-266/Region: protein kinase ATP-binding motif

Alignment Scores:	
Pred. No.:	6.38
Score:	92.00
Percent Similarity:	36.65%
Best Local Similarity:	23.49%
Query Match:	6.67%
DB:	2
	14
	86
Indels:	92
Mismatches:	37
Conservative:	66
Matches:	1146
Length:	

US-09-658-824-808 (1-781) x B35962 (1-1146)

36	AGGTCGTGGATTCTTTCCGCTACTCAGACACGGCGGGTAGGTCCACAGGCAGATCCAAAC	95	QY
613	LYSGlyGlyPhe-PheSerSerPheMetLysLysArgAsnAlaPro---ThrProProly	631	Db
96	TGGAGTTGAAGTGTGAGTCAGAGTCACAGGAACA---GC	134	QY
631	sArgSerSerPheArgGluMetGluAsnGlnProHisLysLysTyrGluLeuThrG	651	Db
135	AGGCTCCGGAGGGTGTG-----TGTCAGTCAGTCACAGATGA	173	QY
651	YAsnPheSerSerValAlaSerLeuGlnHisAlaAspGlyPheSerPheThrProAlaG	671	Db
174	GAAGGCCCTCAAGTCGTGCTCCCTCTCATGCGTGCCAGCCGCAATGGACCTTCTTGCT	233	QY
671	nGlnGluAlaAsnLeuValProPro-----LysCys-----TyrGlyGlySerPheAl	687	Db
234	CGTCAGGCCATAACTAGGAGGAAGAGGGCGGAGGAGCTGGAGGGGCTCAGCGGAAGCT	293	QY
687	aGlnArgAsnLeuCysAsnAspGlyGly-----GlyGlyGlySerGlyThrAl	705	Db
294	GGGTGCTGTGG-----GGGTAT-----	312	QY
705	aGlyGlyGlyTrpSerGlyIleThrGlyPhePheThrProArgLeuIleLysLysThrLe	725	Db
313	-----CCGATGCCACAGAAG	326	QY
725	uGlyLeuArgAlaGlyLysProThrAlaSerAspAspThrSerLysProPheProArgSe	745	Db
327	CACCTGGAAACCCGACAGAGATTCTGGACTCCCCAGACGGGACCAAG-----	373	QY
745	rAsn-SerThrSerSerMetSerSerGlyLeuProGluGlnAspArgMetAlaMetThrL	765	Db
374	-----CAGAGGGACGCCATGACGGCACACAC	398	QY
765	euProArgAsnCysGlnArgSerLysLeuGlnLeuGluArg-Thr-----ValSer	781	Db
399	ACAAACACAGAACACACAGCCAGTCCAGGAGCCCAAGTAATGGAGAGCCCAAAAA---	455	QY
782	ThrSerSerGlnProGluGluAsnValAspArgAlaAsnAspMetLeuProLysLysSer	801	Db
456	GAAGAACCAGCAGCTGAAAGTCGGGATCTACACCTGGGACGACAGACAGAAGAATCAG	515	QY
802	GluGluSerAlaAlaProSerArgGluArgProLysAlaLysLeuLeu-----ProArg	819	Db
516	GATACAGCTCAGATCCAGTCGCGGACATGGGAAGGTGATCTCAAGAGCTGCATCATCA	575	QY
820	GlyAlaThrAlaLeuProLeuArgThrProSerGlyAspLeuAlaIleThrGluLys-As	839	Db
576	AACACCGGGGATAAATCTGGATTTCGGGCTCAAGGTGAACATAATACCTAAAGA	635	QY
839	pProProGlyValGlyVal-----AlaGlyValAlaAlaProLysGlyLysG	856	Db
636	GGAACAC-----TGTAATAATGCCAGACGAGCTGAAGAG	669	QY
856	uLysAsnGlyValAlaArgLeuGlyMetAlaGlyValProGluAspGlyGluGln	874	Db

RESULT 33



A35962

protein-tyrosine kinase (EC 2.7.1.112) I, splice form B - human

N:Alternate names: c-abl proto-oncogene homolog 2

C:Species: Homo sapiens (man)

C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 04-Feb-2000

C:Accession: A35962; A47577

R:Kruh, G.D.; Perreg6, R.; Miki, T.; Aaronson, S.A. Proc. Natl. Acad. Sci. U.S.A. 87, 5802-5806, 1990

A:Title: The complete coding sequence of arg defines the Abelson subfamily of cytoplasmic tyrosine kinases

A:Reference number: A35962; MUID:90332670

A:Accession: A35962

A:Molecule type: mRNA

A:Residues: 1-1182 <KRU1>

A:Cross-references: GB:M35296; NID:g178992; PIDN:AAA35553.1; PID:g178993

R:Kruh, G.D.; King, C.R.; Kraus, M.H.; Popescu, N.C.; Amsbaugh, S.C.; McBride, W.O.; Aar

Science 234, 1545-1548, 1986

A:Title: A novel human gene closely related to the abl proto-oncogene.

A:Reference number: A47577; MUID:87069952

A:Accession: A47577

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: '11', 345-469 <KRU2>

A:Cross-references: GB:M14903; NID:g177954; GB:M14904; NID:g177955

C:Genetics:

A:Gene: GDB:ABL2; ABL1

A:Cross-references: GDB:119641; OMIM:164690

A:Map position: 1q24-1q25

C:Superfamily: human protein-tyrosine kinase abl; protein kinase homology; SH2 homology;

C:Keywords: alternative splicing; ATP; kinase-related transforming protein; phosphotrans

F:114-162/Domain: SH3 homology <SH3>

F:173-263/Domain: SH2 homology <SH2>

F:286-546/Domain: protein kinase homology <KIN>

F:294-302/Region: protein kinase ATP-binding motif

Alignment Scores:

Pred. No.:	6, 4	Length:	1182
Score:	92.00	Matches:	66
Percent Similarity:	36.65%	Conservative:	37
Best Local Similarity:	23.49%	Mismatches:	92
Query Match:	6.67%	Indels:	86
DB:	2	Gaps:	14

US-09-658-824-808 (1-781) x A35962 (1-1182)

QY	36	AGGTCGTGATCTTTCCCGCTACTGAGACACGGCGGTAGGTCACAGCAGATCCAAAC	95
Db	649	LysGlyGlyPhe-PheSerSerPheMetLysLysArgAsnAlaPro---ThrProProLy	667
QY	96	TGGGAGTTGAAGTGTGAGTCAGAGTGAAGGAACCA-----GC	134
Db	667	sArGSerSerPheArgGluMetGluAsnGlnProHisLysLysTyrGluLeuThrG1	687
QY	135	AGGCTCCGAGGGTTGTG-----TGTCAGTCACTCAGAGTGA	173
Db	687	YAsnPheSerSerValAlaSerLeuGlnHisAlaAspGlyPheSerPheThrProAlaG1	707
QY	174	GAAGGCCCTCGAGTCGTGCTCCTCTCATGCGTGCCAGCCCATGGACCTTCTGTGCT	233
Db	707	nGlnGluAlaAsnLeuValProPro-----LysCys-----TyrGlyGlySerPheA1	723
QY	234	CGTCAGGCCATAACTAGGGAAGAGGCGCGAGGAGTGAGGGGCTCAGCGGAAGCT	293
Db	723	aGlnArgAsnLeuCysAsnAspGlyGly-----GlyGlyGlyGlySerGlyThrA1	741
QY	294	GGGTGCTGTGG-----GGGTAT-----	312
Db	741	aGlyGlyGlyTrpSerGlyIleThrGlyPhePheThrProArgLeuIleLysLysThrLe	761
QY	313	-----CCGAGTCCCAAG	326
Db	761	uGlyLeuArgAlaGlyLysProThrAlaSerAspAspThrSerLysProPheProArgSe	781
QY	327	CACCTGGACCCGACAGAAGATTCCTGACTCCCAAGCGGACCAAG-----	373

```
Db 1419 -----ThrArgGlyArgGlnGlySerHisHisGluGlnAla 1430
QY 216 CCATGGACCTTCCTGCTCTCTG-----CACGGC 242
Db 1431 GlnAspSer-----SerArgHisSerAlaSerGlnGluGlyGlnAspThrIleHisGly 1448
QY 243 CATAACTAGGAGGAGGAGGCGGAGGAGTGGAGGGCTCAGCGCAAGCTGGGGTGCTG 302
Db 1449 HisPro---GlyGlyHisSerAlaAspSerSerArgGlnSerGly-----1462
QY 303 TTGGGGGTATCCGAGTCCCAAGACCTCGGAACCCCGACAGAAAGATTCTGGACTCCCCA 362
Db 1463 -----ThrArgHisThrGlu-----1467
QY 363 GACGGGACAGGAGGAGCGGATGAGCGGCACACACACAAACACAGAACCA-----413
Db 1468 ---SerSerSerArgGlyGlnAlaAlaSerHisGluGlnAlaArgSerSerAlaGly 1486
QY 414 -----CACAGCCAGTCCAGGAGCCCAAGT-----GGA 443
Db 1487 GluArgHisGlySerHisHisGlnGlnSerAlaAspSerSerArgHisAlaGlyIleGly 1506
QY 444 GAGCCCCAAAGAAGAACAGCAGCTGAAAGT-----CGGGAT 482
Db 1507 HisGlyGlnAlaSerSerAlaValArgAspSerGlyHisArgGlyTyArgGlySerGln 1526
QY 483 CCTACACCTGGGAGCAGACAGAAAGATCAGGATACAGCTCAGATCCCGAGTGGCGGAC 542
Db 1527 AlaThrAspSerGluGlyHisSerGluAspSerAspThrGlnSerValSerAlaGlnGly 1546
QY 543 ATGGAAGTGTATCGCAAGAGCTGCATCAGTCAACACCGGGGATAAATCTGGATTGGG 602
Db 1547 GlnAlaGlyProHisGlnGlnAlaHisGlnGlnSerAlaArgGlyGlnSerGlyGluSer 1566
QY 603 TTCGGCGGTCAAGGT 617
Db 1567 SerGlyArgSerGly 1571
RESULT 35
B87676
hypothetical protein CC3444 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: B87676
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: B87676
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-268 <STO>
A:Cross-references: GB:AE005673; NID:g13425162; PIDN:AAK25406.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC3444
Alignment Scores:
Pred. No.: 6.36 Length: 268
Score: 91.50 Matches: 42
Percent Similarity: 34.81% Conservative: 13
Best Local Similarity: 26.58% Mismatches: 56
Query Match: 6.63% Indels: 9
DB: 2 Gaps:
US-09-658-824-808 (1-781) x B87676 (1-268)
QY 262 GGCGCAGGAGTGGAGGGCTCAGGGAAGCTGGGTGCTGTGGGGGTAT-----312
Db 12 GlyPro---SerValGlyPheAsnAlaValLeuSerSerCysTrpGlyTyrlLeuAlaTyr 30
QY 313 ---CCGAGTCCAGAAAGCAGCTCGGAACCCCGACAGA-----AGATTCTGGACTCCC 360
```

```
Db 31 AlaProCysAlaArg-----GlnProAspArgThrArgProArgAspTrp-----45
QY 361 CACAGCGGACGACGAGGAGGAGGCATGAGCAGACACACAAACACACAGCCACACAGCC 420
Db 46 -----GlyArgGlyArgPheHisArgArgSerGlyThrAsnArgArgSer 61
QY 421 AGTCCCAGAGCCCACTAATGGAGAGCCCAA-----AAAGAAGAACACAGCAGCTGAAAGT 476
Db 62 GlyProAlaAlaProProAlaGlnAlaProAlaAlaProAlaAlaGlnAla 81
QY 477 CGGGAT-----CCTACACCTGGCAGCAGACAGCAAGAAGATCAG 515
Db 82 AlaProAlaAspGlnAlaAlaProValAlaAlaProAlaAlaGluThrAlaProAlaAla 101
QY 516 GATACAGCTGAGATCCAGTCCGCGCAGCATGGAA-----548
Db 102 AspThrTyrrValAlaProValArgGlyProArgThrThrAspProPheThrValArgLeu 121
QY 549 -----GGTGATCTGCAAGAGCTG 566
Db 122 LeuAspValLeuAspLysValCysLysProGlnValAlaGlyAspPheAlaGlnLeu 141
QY 567 CATCAGCTAAACACCGGGGATAAATCTGGATTGGGTTCGGCGCTCAAGGTGAA 620
Db 142 ValLysSer-----TyrGlyPheLysLysArgGlu 152
RESULT 36
S24169
mucin - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 10-Nov-1995
C:Accession: S24169
R:Huan, L.J.; Xu, G.; Forstner, G.; Forstner, J.
Biochim. Biophys. Acta 1132, 79-82, 1992
A:Title: A serine, threonine and proline-rich region near the carboxyl-terminus of a
A:Reference number: S24169; MUID:92379096
A:Accession: S24169
A:Molecule type: mRNA
A:Residues: 1-292 <HUA>
Alignment Scores:
Pred. No.: 6.4 Length: 292
Score: 91.50 Matches: 39
Percent Similarity: 40.16% Conservative: 10
Best Local Similarity: 31.97% Mismatches: 56
Query Match: 6.42% Indels: 17
DB: 2 Gaps:
US-09-658-824-808 (1-781) x S24169 (1-292)
QY 357 AGTCCAGAAATCTCTCGGGGTTCAGGTGCTTCTGGACTCGGATA---CCCCAACA 301
Db 90 ThrProThrSerThrSerThrValProThrThrGlySerThrSerSerLysProProThr 109
QY 300 GCACCCCACTTCGCCCTG---AGCCCTCCACTCTCGGCCCTCTCTCTCCCTAGTTAT 244
Db 110 GlySerSerThrProIleThrSerProSerThrPro-SerProProThrGlu-----126
QY 243 GGCGGTGACGACAGAAAGGTCCATGGCGTGGCACCAGCATGAGAGGACGACGACTTC 184
Db 127 -SerThrThrLeuSerSerThrPro-----ValThrThrThr 138
QY 183 GAGGGCTTCTCACTCTGAGTCACTGACCACACAAACCCTCCGGAAGCCTGCTGGTTCCTC 124
Db 138 rAlaThrSerSerThrThrSerSerProGlyThrThr-----SerProPheAlaThrSe 156
QY 123 TTCACCTCTCACTCACACTTCAACTCCCACTGGATCGCTGTGGACCTACCCGCGTGT 64
Db 156 rSerValSerSerThrProProSerProSerProSerSerGlyProThrThrSerSe 176
QY 63 CTCAGTAGCGGAGAAAGATCCAGACCTCAGGGACCCGAGTGCAGCGCTCACAGCTCCGC 4
```



R;Yuryev, A.; Patturajan, M.; Litvintung, Y.; Joshi, R.V.; Gentile, C.; Gebara, M.; Cord  
C;Accession: T31422  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A;Residues: 1-1473 <YUR>  
A;Cross-references: EMBL:U49057; NID:g1438533; PID:g1438534; PIDN:AAC52658.1  
A;Experimental source: hippocampus  
A;Reference number: Z21024; MUID:96293459

Search completed: October 18, 2002, 09:45:49  
Job time : 58 secs

RESULT 40  
S70009

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	550	39.9	146	1	GGD2_HUMAN	GG1664	homo sapien
2	138.5	10.0	102	1	GGC1_HUMAN	O60829	homo sapien
3	114	8.3	908	1	SRCA_RABIT	P13666	oryctolagus
4	110	8.0	146	1	GGEL_HUMAN	O73459	homo sapien
5	106	7.7	246	1	GRPA_RAT	P08568	rattus norv
6	106	7.7	527	1	G19P_HUMAN	P14314	homo sapien
7	103.5	7.5	2472	1	NCR2_MOUSE	O9wu42	mus musculu
8	103	7.5	247	1	GRPB_RAT	P08462	rattus norv
9	103	7.2	363	1	CYB_TRYBB	P00164	trypanosoma
10	99.5	7.2	503	1	DMP1_MOUSE	O55188	mus musculu
11	99.5	7.0	598	1	NR41_HUMAN	P22736	homo sapien
12	98	7.1	1736	1	CA2B_HUMAN	P13942	homo sapien
13	97	6.8	371	1	CYB8_LEITA	P14548	leishmania
14	96.5	7.0	277	1	HMPH_CHICK	Q05502	gallus gall
15	96	6.7	348	1	Y735_STRPR	P20186	streptomyce
16	95.5	6.9	117	1	GGE4_HUMAN	Q13068	homo sapien
17	95.5	6.9	117	1	GGE5_HUMAN	Q13069	homo sapien
18	95.5	6.9	117	1	GGE6_HUMAN	Q13070	homo sapien

```
Best Local Similarity: 52.19% Mismatches: 1
Query Match: 39.86% Indels: 108
DB: 1 Gaps: 1

US-09-658-824-808 (1-781) x GGD2_HUMAN (1-146)

QY 2 CGGCGGAGCTGTGAGCGCGGCGACTCGGTCCTGAGGTGCTGATCTTTCCTCGCTACTG 61
DB 27 ATGAGS22CysGluProAlaThrArgValProGluValTrpIleLeuSerProLeuLeu 46
QY 62 AGACAGCGGGGTAGTCCACAGGCGAGATCCAACCTGGGAGTTGAAGTGTGAGTGAGAGTG 121
DB 47 ArgHisGly----- 49
QY 122 AAGAGGAACACAGAGGCTTCGCGAGGGTTGTGTGTCAGTCACTCAGAGTGAGAGGCGCC 181
DB 49 ----- 49
QY 182 TCGAAGTCGTCCTCCTCTCTCATCGGTGCCGCCCATCGACCTTCTTGTCTGTCTGTCACGG 241
DB 49 ----- 49
QY 242 CCATAACTAGGAGGAGGCGGCGGAGTGAGGGGCTCAGCGGAAGCTGGGGTGGT 301
DB 49 ----- 49
QY 302 GTTGGGGGTATCCGAGTCCAGAGCAACCTCGAACCCTCGACAGAGATTCGTGACTCCCC 361
DB 49 ----- 49
QY 362 AGACGGGACGAGGAGGCGGCGCATGAGCGGACACACACAAACACAGAACACACAGCCA 421
DB 50 -----Gly--HisThrGlnAsnHisThrAlaAs 60
QY 422 GTCCAGGAGCCCACTAATGAGAGCCCAAAAGAGAACACAGCAGCTGAAAGTCGGGA 481
DB 60 erProArgSerProValMetGluSerProLysLysLysAsnGlnGlnLeuLysValGlyI 80
QY 482 TCCTACACTGGGACGACAGAGAGATCAGGATCAGCTGAGATCCAGTCGCGGGA 541
DB 80 leLeuHisLeuGlySerArgGlnLysLysIleArgIleGlnLeuArgSerGlnCysAla 100
QY 542 CATGAAGGTGATCTGCAAGAGCTGCATCAGTCAAAACACCGGGGATAAATCTGATTGG 601
DB 100 hrTriLysValIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeu 120
QY 602 GTTCCGGCTCAAGTGAAGATATACCTAAGAGGAACACTGTAAATGCCAAGACAG 661
DB 120 lyserGlyValLysValLysIleIleProLysGluGluHisCysLysMetProGluAla 140
QY 662 GTGAAGAGCAACCAAGTT 681
DB 140 lyGluGluGlnProGlnVal 146

RESULT 2
GGC1_HUMAN
ID GGC1_HUMAN STANDARD; PRT; 102 AA.
AC O60829;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE G antigen family C.1 protein (Prostate-associated gene protein 4)
DE (PAGE-4) (PAGE-1) (JM27) (GAGE-9).
GN GAGE1 OR PAGE4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98393718; PubMed=9724777;
RA Brinkmann U., Vasmatazis G., Lee B., Yerushalmi N., Essand M.,
RA Pastan I.;

*PAGE-1, an X chromosome-linked GAGE-like gene that is expressed in
normal and neoplastic prostate, testis, and uterus.*;
Proc. Natl. Acad. Sci. U.S.A. 95:10757-10762(1998).
[2]
RP SEQUENCE FROM N.A.
RA Strom T.M., Nyakatura G., Hellebrand H., Drescher B., Rosenthal A.,
RA Weindl A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: UNKNOWN.
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN NORMAL MALE AND
CC FEMALE REPRODUCTIVE TISSUES, PROSTATE, TESTIS, FALLOPIAN TUBE,
CC UTERUS, AND PLACENTA, AS WELL AS IN PROSTATE CANCER, TESTICULAR
CC CANCER, AND UTERINE CANCER.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF275258; AAF88037.1; -
CC EMBL; AJ005894; CAA06751.1; -
CC EMBL; AF238380; AAF62541.1; -
CC MIN: 300287; -
CC
CC SEQUENCE 102 AA; 11153 MW; CE5D07AFBF73301B CRC64;

Alignment Scores:
Pred. No.: 0.00201 Length: 102
Score: 138.50 Matches: 34
Percent Similarity: 55.70% Conservative: 10
Best Local Similarity: 43.04% Mismatches: 31
Query Match: 10.04% Indels: 4
DB: 1 Gaps: 1

US-09-658-824-808 (1-781) x GGC1_HUMAN (1-102)

QY 441 GGAGAGCCCCAAAAGAACAGACAGCAGCTGAAAGTCGGGATCTACACCTGGCGCAG 500
DB 28 GlyGluSerGlnGlnGluGluProThrAspAsnGlnAspIleGluProGlyGlnGlu 47
QY 501 ACAGAAGAATCAGGATACAGCTGAGATCCAGTCGCGCAGCATGGAAGTGTGATCGCA 560
DB 48 ArgGlu-----GlyThrProIleGluGluArgLysValGluGlyAspCysGln 64
QY 561 GAGCTGCATCAGTCAACACACCGGGGATAAATCTCGATTGGTTCGCGCTCAAGTGAA 620
DB 65 GluMetAspLeuGluLysThrArgSerGluArgGlyAsp-GlySerAspValLysGluLy 84
QY 621 GATAATACCTAAGAGGAACACACTGTAAATGCCAGAGCAGGTGAACAGCAACCA 675
DB 84 sThrProAsnProLysHisAlaLysThrLysGluAlaGlyAspGlyGlnPro 102

RESULT 3
SRCRABIT
ID SRCRABIT STANDARD; PRT; 908 AA.
AC P13666;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sarcalumenin precursor.
DE Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89345602; PubMed=2762314;
RA Leberer E., Charuk J.H.M., Green N.M., MacLennan D.H.;
RA "Molecular cloning and expression of cDNA encoding a luminal calcium
RT binding glv opote'n from sarcoplasmic reticulum.";
```

QY	340	GACACAAGATTCTGCAGTCC	CCACACGGGACGAGGAGCGGATGAGCGACACACA	399
		:::     :::	:::	
Db	394	rSerGluGluAlaGlyAlaAlaSer	GluGluAlaSer-GlyThrAlaGlyLeuGlyGluG	414
QY	400	CAAACACAGAACCCAC	-----AGCCACTCCGAGG	429
			:::	
Db	414	luGluThrGlnProSerThr	GluGlyLeuAspSerGlyProAlaGlySerGlnAlaGlnA	434
QY	430	ACCCAGTAATGGAGAGCC	CAAAAGAACAGACAGCTGAAGCTCGGATCTCTACAC	489
		:::     :::		
Db	434	spThrGluAlaGluProGluGluGlyHisGlnGly	ProGluGlyProIleThrAlaLap	454
QY	490	CTGGCAGCAGACAGAACAGATCAG	TACAGCTGAGATCCACGTCGCGCAGCATGGAAG	549
		:::     :::	:::	
Db	454	roGlnGluThrGluaspValSerGlu	-----GluValProMetArgasp	469
QY	550	GTGATCTGCAAGAGCTGCATCAGT	CAAACACCGGGGATAATCTGGATTTGGGTTCCGGC	609
		469	-----	469
QY	610	GTCAAGGTCAAGATATACCTAA	AGAGCAACTGTAAATGCCAAGACGAGGTGAAGAG	669
			:::	
Db	470	-----	-----ArgSerHisIleGlnL	475
QY	670	CAACCACAAGTTTAAATGAAGACA	AG 695	
Db	475	ysThrLeuMetLeuaspGluaspLys	483	
RESULT 4				
GGBI_HUMAN				
ID	GGBI_HUMAN	STANDARD;	PRT; 146 AA.	
IC	O75459;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	G antigen family B 1 protein (prostate-associated gene protein 1)			
DE	(PAGE-1) (GAGE-9) (AL5).			
GN	GAGEB1 C <sub>9</sub> PAGE1 OR GAGE9.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid:9606;			
RP	[1]			
RN	SEQUENCE FROM N.A.			
RA	MEDLINE-98316329; PubMed-9651357;			
RX	Chen M.E., Lin S.-H., Chung L.W.K., Sikes R.A.;			
RT	*Isolation and characterization of PAGE-1 and GAGE-7; new genes			
RT	expressed in the LNCaP prostate cancer progression model that share			
RT	homology with melanoma-associated antigens.;			
RL	J. Biol. Chem. 273:17618-17625(1998).			
CC	-1- FUNCTION: UNKNOWN.			
CC	-1- TISSUE SPECIFICITY: ISOLATED FROM PROSTATE CANCER CELL LINES;			
CC	EXPRESSION ASSOCIATED WITH PROGRESSION TO ANDROGEN INSENSITIVE			
CC	PHENOTYPE. EXPRESSED IN NORMAL TESTIS AND AT LOWER LEVEL IN NORMAL			
CC	PLACENTA.			
CC	-1- SIMILARITY: BELONGS TO THE GAGE FAMILY.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collabora			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstati			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no			
CC	modified and this statement is not removed. Usage by and for commerc			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/annou">http://www.isb-sib.ch/annou</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC				
DR	EMBL; AF058989; AAC25990.1; -			
DR	MTM; 300288; -			
KW	Multigene family.			
SQ	SEQUENCE 146 AA; 16134 MW; 91ABA1E3B498DCA2 CRC64;			
Alignment Scores:				
Pred. No.:		0.329	Length:	146









Db 918 AlaThrCysSerAlaAspGluValAspGluProGluGlyGlyAspLys----- 933

QY 603 TTCCGGCGTCAAGTGAAGATAATACCTAAAGAGGACACTGTAAATGCCAGAGCAGG 662

Db 934 -----GlyArgLeuLeuSerProArgProSerLeuLeuThrProAlaGlyAsp 949

QY 663 TGAAGACCAACCACT-----TTAAATCAAGACACAGCTGAACACACGC 707

Db 950 ProArgAlaSerThrProGlnLysProLeuAspLysGlnLeuLysGlnArg 968

RESULT 8

GRPB\_RAT

ID GRPB\_RAT STANDARD; PRT; 247 AA.

AC P08462;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Submandibular gland secretory GLX-rich protein CB precursor (GRP-CB)

DE (Contiguous repeat polypeptide) (CRP).

GN GRP-CB.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=87165965; PubMed=3558393;

RA Heinrich G., Habener J.F.;

RT \*Genes encoding proteins with homologous contiguous repeat sequences

RT are highly expressed in the serous cells of the rat submandibular

RT gland.\*;

RL J. Biol. Chem. 262:5262-5270(1987).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Wistar;

RX MEDLINE=91139639; PubMed=1995617;

RA Cooper L.F., Elia D.M., Tabak L.A.;

RT \*Secretagogue-coupled changes in the expression of glutamine/glutamic

RT acid-rich proteins (GRPs). Isoproterenol induces changes in GRP

RT transcript expression and changes in isoforms secreted.\*;

RL J. Biol. Chem. 266:3532-3539(1991).

CC -1- FUNCTION: GRP PROTEINS HAVE A MARKED AFFINITY FOR HYDROXYAPATITE.

CC THEY MAY PLAY A ROLE IN THE FORMATION OF THE PROTECTIVE ACQUIRED

CC PELLICLE AT THE SALIVA-TOOTH INTERFACE.

CC -1- SUBCELLULAR LOCATION: Extracellular.

CC -1- TISSUE SPECIFICITY: SUBMANDIBULAR GLAND ACINAR CELLS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M31032; AAA40969.1; -

DR EMBL; M58654; AAA41279.1; ALT\_SEQ.

DR PIR; A29545; A29545.

DR PIR; B38647; B38647.

KW Submandibular gland; Saliva; Repeat; Signal; Multigene family.

FT SIGNAL

FT CHAIN 1 18

FT CHAIN 19 247

FT DOMAIN 67 181

FT REPEAT 67 89

FT REPEAT 90 112

FT REPEAT 113 135

FT REPEAT 136 158

FT REPEAT 159 181

FT REPEAT 1 207

FT IDENTICAL TO RAT GRP-CA (P08568).

SEQUENCE 247 AA; 26948 MW; 72A5075BCCC314E4 CRC64;

Alignment Scores:

Pred. No.: 1.19 Length: 247

Score: 103.00 Matches: 39

Percent Similarity: 40.29% Conservative: 17

Best Local Similarity: 28.06% Mismatches: 50

Query Match: 7.46% Indels: 33

DB: 1 Caps: 6

US-09-658-824-808 (1-781) x GRPB\_RAT (1-247)

QY 306 GGGGTATCCGAGTCCCGAAGACACCTGTAACCGAGACAGAGATTCTGGACTCCCGAC 365

Db 74 GlySerGluGluGluGlnGlnGlnGluProThrGlnAlaGluAsnGlnGluPro 93

QY 366 GGGACAGAGAGAGGCGCATGAGCGACACACACAACACACACAGCCAGTC 425

Db 94 AlaThr-----SerGlySerGluGluGlnGlnGln 104

QY 426 CAGGAGCCCACT-----AATGGAGAGCCC----- 449

Db 105 GlnGluProThrGlnAlaGluAsnGlnGluProProAlaThrSerGlySerGluGlu 124

QY 450 ---CAAAAGAAGAACCAACGCA---GCTGAAAGTCGGATCTACACCTGGGCAGCAGCA 503

Db 125 GlnGlnGlnGlnGlnProThrGlnAlaGluAsnGlnGluProProAlaThrSerGlySer 144

QY 504 GAAGAATCAGGATACAGCTGAGATCCAGTCGCGACATGGAAGTGATCTGCAAGAG 563

Db 145 GluGluGluGlnGlnGlnGlu-----SerThrGlnAlaGluAsnGlnGlu 160

QY 564 CTGCATCAGTCMAACACCGGGGATAATCTGGATTGGTTGGGTCGCGCTCAAGGTGAAGAT 623

Db 161 -----ProSerAspSerAlaGlyGlnGlnGlnGlnProGluGlu 175

QY 624 AATACCTAAAGAGGACACACTGTAAATGCCAGAGCAGGTGAAGACGACCAACAA 678

Db 176 GlyAsn-ValGluSerProProSerProGluAsnSerGlnGlnGlnProGln 193

RESULT 9

CYB\_TRYBB

ID CYB\_TRYBB STANDARD; PRT; 363 AA.

AC P00164;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Cytochrome B.

GN COB OR CYTB.

OS Trypanosoma brucei brucei.

OG Mitochondrion.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

OX NCBI\_TaxID=5702;

RN [1]

RP SEQUENCE FROM N.A. (KINETOPLAST).

RX MEDLINE=84041494; PubMed=6314266;

RA Benne R., de Vries B.F., van den Burg J., Klaver B.;

RT \*The nucleotide sequence of a segment of Trypanosoma brucei

RT apocytchrome b and some unusual unassigned reading frames.\*;

RL Nucleic Acids Res. 11:6925-6941(1983).

CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE

CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A

CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL

CC COUPLED TO ATP SYNTHESIS.

CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY

CC BOUND TO THE PROTEIN.

CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,

CC CYTOCHROME C1 AND THE RIESKE PROTEIN.

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way



```

QY 142 CGGAGGTTGTGTGTCAGTCACTGAGAGCCCTCGAAGTCGTGCTCCCTCTC 201
Db 138 -----AspThrThrGlnSerSerGlu-AspSerThrSerGlnGluAsnSe 152
QY 202 ATCGGTGCCAGCCCATGGACTCTCT-----GTCCTGCTCAGC 240
Db 152 rAlaGlnAspThrProSerAspSerLysAspGlnAspSerGluAspAspAlaHisSerAr 172
QY 241 GCCATACTAGG-----AGGAAGGAGCCGAGG 270
Db 172 gProAspAlaGlyAspSerAlaGlnHisSerGluSerGluGlnArgValGlyGyl 192
QY 271 AGTGGAGGGCTCAGCGAAGCTGGGT----- 298
Db 192 ySerGluGly---GlnSerSerHisGlyAspGlySerGluPheAspAspGluGlyMetGl 211
QY 299 -----CCTGTTGGGG 309
Db 211 nSerAspAspProGluSerThrArgSerAspArgGlyHisAlaArgMetSerSerAlaGl 231
QY 310 TATCCGAGTCCCAAGACACCTGGA-----ACCCGACAGAAAGATTCGGACT 357
Db 231 ylleArgSerGluSerLysGlyAspArgGluProThrSerThrGlnAspSerAspAs 251
QY 358 CCCACAG----- 364
Db 251 pSerGlnSerValGluPheSerSerArgLysSerPheArgSerArgValSerGluGl 271
QY 365 -----CGGACACAGGAGGAGCGCATGAG 390
Db 271 uAspTyrArgGlyGluLeuThrAspSerAsnSerArgGluThrGlnSerAspSerThrGl 291
QY 391 CGACACA---CACAAACACAGAACACACAGCCAGTCCAGGAGCCCAATATGGAGAC 447
Db 291 uAspThrAlaSerLysGlu-GluSerArgSerGluSerGlnGlu---AspThrAlaGlu 310
QY 448 CCATAAAGAACACCGAGCTGAAGTCGGATCTACACCTGGCGGACGACAGAGAAG 507
Db 310 erGlnSerGlnGluAspSerProGluGlyGlnAspProSerSerGluSerSerGluGlu 330
QY 508 AAGATCAGGATACAGCTGAGATCCAGTCCGACATGGAAGTGTATCTGCAAGAGCTGC 567
Db 330 laGlyGluProSerGlnGluSerSerSerGluSerGlnGluGlyValThrSerGluSerA 350
QY 568 ATCATGCAACACCGGGGATAATCTGGATTGGTTCGGCGCTCAAGGTGAAGATAAT 626
Db 350 rgGlyAspAsnProAspAsnThrSerGlnThrGlyAspGlnGluAspSerGluSerSer 369

RESULT 11
NR41_HUMAN STANDARD; PRT; 598 AA.
ID P22736;
AC 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Orphan nuclear receptor HMR (Early response protein NAK1) (TR3 orphan
DE receptor).
GN NR41 OR HMR OR NAK1 OR GFRP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=fetal skeletal muscle;
RX MEDLINE=91133413; PubMed=2283997;
RA Nakai A., Kartha S., Sakurai A., Toback F.G., Degroot L.J.;
RT "A human early response gene homologous to murine nur77 and rat
RL NGFI-B, and related to the nuclear receptor superfamily.*;
RN Mol. Endocrinol. 4:1438-1443(1990).
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=90173205; PubMed=2626032;
RA Chang C., Kokontis J., Liao S., Chang Y.;
RT "Isolation and characterization of human TR3 receptor: a member of
RT steroid receptor superfamily.*;
RL J. Steroid Biochem. 34:391-395(1989).
CC -1- FUNCTION: ORPHAN NUCLEAR RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: FETAL MUSCLE AND ADULT LIVER, BRAIN, AND
CC THYROID.
CC -1- INDUCTION: BY GROWTH-STIMULATING AGENTS.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR4 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D49728; BRA08565.1; -.
DR EMBL; L13740; AAA36763.1; -.
DR PIR; A37251; A37251.
DR HSSP; P19793; 2NULL.
DR TRANSFAC; T02767; -.
DR MIM; 139139; -.
DR InterPro; IPR000536; Hormone_rec_llg.
DR InterPro; IPR001723; Strdhormone_receptor.
DR InterPro; IPR001628; zf-C4.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; ZnF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Phosphorylation.
FT DNA_BIND 267 332 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 267 287 C4-TYPE.
FT ZN_FING 303 327 C4-TYPE.
FT DOMAIN 409 459 LIGAND-BINDING (POTENTIAL).
FT DOMAIN 82 92 POLY-SER.
FT DOMAIN 583 586 POLY-PRO.
FT MOD_RES 341 341 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT MOD_RES 351 351 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT CONFLICT 262 262 G -> P (IN REF. 2).
FT CONFLICT 370 370 R -> L (IN REF. 2).
SQ SEQUENCE 598 AA; 64463 MW; 41DAEA7C25FDA22 CRC64;

Alignment Scores:
Pred. No.: 2.41 Length: 598
Score: 99.50 Matches: 56
Percent Similarity: 37.11% Conservative: 16
Best Local Similarity: 28.87% Mismatches: 72
Query Match: 6.98% Indels: 50
DB: 1 Gaps: 7

US-09-658-824-808 (1-781) x NR41_HUMAN (1-598)
QY 568 TGCAGCTCTTGCAGATCACTTCATGTCGCGACATGGGATCTACGCTATCTCCTGATCT 509
Db 78 CysSerSerAlaSerSerAlaSerSerThrSerSerSerAlaThrSerProAla 97
QY 508 TCTTCTGTC-----TGCTGCCAGCT----- 488
Db 98 SerAlaSerPheLysPheGluAspPheGlnValTyrGlyCysTyrProGlyProLeuSer 117
QY 487 -----GTAGGATCCCGACTTCAGCTGCTGGTCTCTCTTTTGGGGCTCTCCATTA--- 437
Db 118 GlyProValAspGluAlaLeuSerSerGlySerAspTyrTyrGlySerProCysSer 137

```





"Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons.";  
Cell 53:401-411(1988).  
(2)  
SEQUENCE FROM N.A.  
MEDLINE=85297791; PubMed=2994021;  
Simpson A.G., Neckelmann N., la Cruz V.F., Muhich M.L., Simpson L.;  
"Mapping and 5' end determination of kinetoplast maxicircle gene transcripts from Leishmania tarentolae.";  
Nucleic Acids Res. 13:5977-5993(1985).  
(3)  
SEQUENCE OF 1-48 FROM N.A.  
MEDLINE=88124876; PubMed=2448777;  
Feagin J.E., Shaw J.M., Simpson L., Stuart K.;  
"Creation of AUG initiation codons by addition of uridines within cytochrome b transcripts of kinetoplastids.";  
Proc. Natl. Acad. Sci. U.S.A. 85:539-543(1988).  
(4)  
SEQUENCE OF 21-371 FROM N.A.  
MEDLINE=85079995; PubMed=6096360;  
de la Cruz V.F., Neckelmann N., Simpson L.;  
"Sequences of six genes and several open reading frames in the kinetoplast maxicircle DNA of Leishmania tarentolae.";  
J. Biol. Chem. 259:15136-15147(1984).  
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL COUPLED TO ATP SYNTHESIS.  
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN.  
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B, CYTOCHROME C1 AND THE RIESKE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
CC -1- CAUTION: THE GENOMIC DERIVED SEQUENCE DIFFERS FROM THAT SHOWN AS IT IS MODIFIED BY EXTENSIVE RNA EDITING.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

[illegible]

Alignment Scores:		
Pred. No.:	3,6	Length: 371
Score:	97.00	Matches: 36
Percent Similarity:	39.47%	Conservative: 24
Best Local Similarity:	23.68%	Mismatches: 46
Query Match:	6.80%	Indels: 46
DB:	1	Gaps: 8
US-09-658-824-808 (1-781) x	CYP-LEITA (1-371)	

QY	705	GTGTGTTTCACCTTGCTCTTCATTTAAACCTTGCTGGTTCCTTCACCTGCCTTC	-----	655
		:::		
Db	45	ileCysGlyValCysLeu-----AlatrPLeuPheSerCysPheIleCysThr	61	
QY	654	---TGCGATTATTCACAGTGTTCCTCTTTAGGTATTATCTTCACCTTGACGCGGAAACCCAA	598	
		:		
Db	62	AsnTrpTrpPheValLeuPheLeuTrpAspPhe-----AspLeuGlyPheVal	77	
QY	597	ATCCAGATTATATCCCGGTTGTTGACGATGACGACTTTCGCATGATCACCTTC	546	
		:		
Db	78	IleArgSerThrHisIleCysPheThrSerLeu-LeuPhePheLeuLeuTrpValHisIle	97	
QY	545	-----CATGTCCGCGACTGGGATCTCAG	523	
		:		
Db	97	ePheLysCysIleValLeuIleLeuPheAspThrHisIleLeuValTrpAlaValGle	117	
QY	522	CTGTATCTCTGATCTCTCTGTCTGTCGCCAGGT----GTAGGATCCCGACTTTCACGCTG	466	
		:		
Db	117	yPheIleIleIleTrpIlePhe-IleValValIleGlyPheIleGlyTrpValLeuProCyst	137	
QY	465	CTGGTCTCTCTTTTGGGGC-----TCTCCATTAC	436	
		:	:	
Db	137	hrMetMetSerTrpTrpGlyLeuThrValPheSerAsnIleLeuAlaThrValProValle	157	
QY	435	TGGGCTCTCTG-----GACTGGCTGTGTGCTCTGTGTTGTG-----T	397	
		:	:	
Db	157	leGlyThrTrpLeuCysTrpTrpIleTrpGlySerGlyTrpIleAsnAspPheThrLeuL	177	
QY	396	GTGTGCGTCATCGCGTCCCTCTCTCTGTCGCCG	365	
			:    :	
Db	177	eulysLeuHisValLeuHisValLeuLeuPro	187	
RESULT 14				
HMHP	CHICK			
ID	HMHP_CHICK	STANDARD;	PRT;	277 AA.
AC	Q05502;			
DT	01-JUN-1994	(Rel. 29, Created)		
DT	01-JUN-1994	(Rel. 29, Last sequence update)		
DT	01-NOV-1995	(Rel. 32, Last annotation update)		
DE	Homeobox protein PRH.			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
EX	MEDLINE=93087175; PubMed=1360645;			
RA	Crompton M.R., Bartlett T.J., Macgregor A.D., Manfioletti G.,			
RA	Burrati E., Giancotti V., Goodwin G.H.;			
RT	"Identification of a novel vertebrate homeobox gene expressed in			
RT	haematopoietic cells.;"			
RL	Nucleic Acids Res. 20:5661-5667(1992).			
CC	-!- FUNCTION: RECOGNIZES THE DNA SEQUENCE 5'-ATTAA-3'. MAY PLAY A ROLE			
CC	IN HEMATOPOIETIC DIFFERENTIATION.			
CC	-!- SUBCELLULAR LOCATION: Nuclear (Probable).			
CC	-!- TISSUE SPECIFICITY: IN ALL HEMATOPOIETIC TISSUES EXCEPT			
CC	PERIPHERAL BLOOD ERYTHROCYTES AND IN THE LIVER AND LUNG.			
-----				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by, and for commercial			
CC	entities requires a license agreement. See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .			
-----				
DR	EMBL; X64711; CAA45966.1; -			
DR	HSSP; P22808; INK3.			
DR	TRANSFAC; T02091; -			
DR	InterPro; IPR001356; Homeobox.			
DR	Pfam; PF00046; homeobox; 1.			
DR	PRINTS; PR00024; HOMEBOX.			



[illegible]

```

RT cytolytic T lymphocytes on a human melanoma.*;
RL J. Exp. Med. 182:689-698(1995).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19145; AA82747.1; .
DR MIM; 604246; .
KW Multigene family.
SQ SEQUENCE 117 AA; 12956 MW; 365ED71B2F9DC7AD CRC64;

Alignment Scores:
Pred. No.: 4.24 Length: 117
Score: 95.50 Matches: 30
Percent Similarity: 45.56% Conservative: 11
Best Local Similarity: 33.33% Mismatches: 47
Query Match: 6.92% Indels: 2
DB: 1 Gaps: 1

US-09-658-824-808 (1-781) x GGE4_HUMAN (1-117)
QY 411 CCACACAGCCAGTCCCGAGGCCAGTAAATGAGAGCCCAAAAGAAAGAACCCAGCAGCT 470
Db 29 ProGluGlnPheSerAspGluValGluProAlaThrProGluGluGlyGluProAlaThr 48
QY 471 GAAAGTCGGGATCTACACCTGGCGAGCAGACAGAGAGATCAGGATACAGCTGAGATC 530
Db 49 GlnArgGlnAspProAlaAlaAlaGlnGluGlyGluAspGluGlyAlaSerAlaGlyGln 68
QY 531 CCAGTCGGCGACATGGAAGTGATCTGCAAGAGCTGCATCAGTCAACACCGGGGATAAA 590
Db 69 GlyProLysPro--GluAlaAspSerGlnGlnGlyHisProGlnThrGlyCysGlu 87
QY 591 TCTGGATTGGGTTCCGCGCTCAAGGTGAAGATATACCTAAAGAGGAACACTGTAAAT 650
Db 88 CysGluAspGlyProAspGlyGlnGluMetAspPro-ProAsnProGluGluValLysTh 107
QY 651 GCACAGACAGCTGAAGAGCAACCAAA 678
Db 107 rProGluGluGlyGluLysGlnSerGln 116

RESULT 17
GGE5_HUMAN
ID GGE5_HUMAN STANDARD; PRT; 117 AA.
AC Q13069;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GAGE-5 protein (G antigen 5).
GN GAGE5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Melanoma;
RX MEDLINE=95378788; PubMed=7544395;
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA Boon T.;
RT "A new family of genes coding for an antigen recognized by autologous
RT cytolytic T lymphocytes on a human melanoma.*;
RL J. Exp. Med. 182:689-698(1995).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -----

```

```

CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19146; AAA82748.1; .
DR MIM; 604247; .
KW Multigene family.
SQ SEQUENCE 117 AA; 12924 MW; 234A865E2ECDD06 CRC64;

Alignment Scores:
Pred. No.: 4.24 Length: 117
Score: 95.50 Matches: 30
Percent Similarity: 45.56% Conservative: 11
Best Local Similarity: 33.33% Mismatches: 47
Query Match: 6.92% Indels: 2
DB: 1 Gaps: 1

US-09-658-824-808 (1-781) x GGE5_HUMAN (1-117)
QY 411 CCACACAGCCAGTCCCGAGGCCAGTAAATGAGAGCCCAAAAGAAAGAACCCAGCAGCT 470
Db 29 ProGluGlnPheSerAspGluValGluProAlaThrProGluGluGlyGluProAlaThr 48
QY 471 GAAAGTCGGGATCTACACCTGGCGAGCAGACAGAGAGATCAGGATACAGCTGAGATC 530
Db 49 GlnArgGlnAspProAlaAlaAlaGlnGluGlyGluAspGluGlyAlaSerAlaGlyGln 68
QY 531 CCAGTCGGCGACATGGAAGTGATCTGCAAGAGCTGCATCAGTCAACACCGGGGATAAA 590
Db 69 GlyProLysPro--GluAlaAspSerGlnGlnGlyHisProGlnThrGlyCysGlu 87
QY 591 TCTGGATTGGGTTCCGCGCTCAAGGTGAAGATATACCTAAAGAGGAACACTGTAAAT 650
Db 88 CysGluAspGlyProAspGlyGlnGluMetAspPro-ProAsnProGluGluValLysTh 107
QY 651 GCACAGACAGCTGAAGAGCAACCAAA 678
Db 107 rProGluGluGlyGluLysGlnSerGln 116

RESULT 18
GGE6_HUMAN
ID GGE6_HUMAN STANDARD; PRT; 117 AA.
AC Q13070;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GAGE-6 protein (G antigen 6).
GN GAGE6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Melanoma;
RX MEDLINE=95378788; PubMed=7544395;
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA Boon T.;
RT "A new family of genes coding for an antigen recognized by autologous
RT cytolytic T lymphocytes on a human melanoma.*;
RL J. Exp. Med. 182:689-698(1995).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U19147; AA82749.1; -  
DR MIM; 604248; -  
KW Multigene family.  
SQ SEQUENCE 117 AA; 12892 MW; 234A865E3FCCCD06 CRC64;

Alignment Scores:  
Pred. No.: 4.24 Length: 117  
Score: 95.50 Matches: 30  
Percent Similarity: 45.56% Conservative: 11  
Best Local Similarity: 33.33% Mismatches: 47  
Query Match: 6.92% Indels: 2  
DB: 1 Gaps: 1

US-09-658-824-808 (1-781) x GGE6\_HUMAN (1-117)

QY 411 CCACAGCCAGTCCAGGAGCCCAAGTAAATGGAGAGCCCAAAAGAAAGAACCCAGCAGCT 470  
DB 29 ProGluGlnPheSerAspGluValGluProAlaThrProGluGluGlyGluProAlaThr 48  
QY 471 GAAAGTCGGATCTACACCTGGCGCAGCAGACAGAAAGATCAGGATACGATCAGATC 530  
DB 49 GlnArgGlnAspProAlaAlaGlnGlnGluGluGlyGlyAlaSerAlaGlyGln 68  
QY 531 CCAGTCGGCAGATCGAAGGTGATCTCGAAGAGCTGCATCAACACCCGGGGATAA 590  
DB 69 GlyProLysPro---GluAlaAspSerGlnGlnGlnGlnHisProGlnThrGlyCysGlu 87  
QY 591 TCTGGATTGGTTCGGCGCTCAAGGTGAAGATTAATACCTAAAGAGGAACACTGTAAT 650  
DB 88 CysGluAspGlyProAspGlyGlnGluValAspPro-ProAsnProGluGluValLysTh 107  
QY 651 GCCAAGACGAGTGAAGACCAACCAAA 678  
DB 107 rProGluGluGlyLysGlnSerGln 116

RESULT 19  
GGE3\_HUMAN STANDARD; PRT; 118 AA.  
AC Q13067;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE GAGE-3 protein (G antigen 3).  
GN GAGE3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Melanoma;  
RX MEDLINE=95378788; PubMed=7544395;  
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,  
Boon T.;  
RT "A new family of genes coding for an antigen recognized by autologous  
cytolytic T lymphocytes on a human melanoma.";  
J. Exp. Med. 182:689-698(1995).  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT  
NOT IN NORMAL TISSUES, EXCEPT TESTIS.  
CC -1- SIMILARITY: BELONGS TO THE GAGE FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U19144; AA82746.1; -  
DR MIM; 604245; -  
KW Multigene family.  
SQ SEQUENCE 118 AA; 12937 MW; D97EEB19E735103 CRC64;

Alignment Scores:  
Pred. No.: 4.25 Length: 118  
Score: 95.50 Matches: 30  
Percent Similarity: 45.56% Conservative: 11  
Best Local Similarity: 33.33% Mismatches: 47  
Query Match: 6.92% Indels: 2  
DB: 1 Gaps: 1

US-09-658-824-808 (1-781) x GGE3\_HUMAN (1-118)

QY 411 CCACAGCCAGTCCAGGAGCCCAAGTAAATGGAGAGCCCAAAAGAAAGAACCCAGCAGCT 470  
DB 30 ProGluGlnPheSerAspGluValGluProAlaThrProGluGluGlyGluProAlaThr 49  
QY 471 GAAAGTCGGATCTACACCTGGCGCAGCAGACAGAAAGATCAGGATACGATCAGATC 530  
DB 50 GlnArgGlnAspProAlaAlaGlnGlnGluGluGlyGlyAlaSerAlaGlyGln 69  
QY 531 CCAGTCGGCAGATCGAAGGTGATCTCGAAGAGCTGCATCAACACCCGGGGATAA 590  
DB 70 GlyProLysPro---GluAlaAspSerGlnGlnGlnGlnHisProGlnThrGlyCysGlu 88  
QY 591 TCTGGATTGGTTCGGCGCTCAAGGTGAAGATTAATACCTAAAGAGGAACACTGTAAT 650  
DB 89 CysGluAspGlyProAspGlyGlnGluMetAspPro-ProAsnProGluGluValLysTh 108  
QY 651 GCCAAGACGAGTGAAGACCAACCAAA 678  
DB 108 rProGluGluGlyLysGlnSerGln 117

RESULT 20  
MY16\_MOUSE STANDARD; PRT; 657 AA.  
AC P17564;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Myeloid differentiation primary response protein MyD116.  
GN MYD116.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SL;  
RX MEDLINE=90251472; PubMed=2339071;  
RA Lord K.A., Hoffman-Liebermann B., Liebermann D.A.;  
RT "Sequence of MyD116 cDNA: a novel myeloid differentiation primary  
Nucleic Acids Res. 18:2823-2823(1990).  
CC -1- INDUCTION: BY INTERLEUKIN-6.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X51829; CAA36128.1; -  
DR PIR; S10001; S10001.  
DR MGD; MGI:1100516; Myd116.  
KW Differentiation; Antigen; Repeat.  
FT DOMAIN 283 451 4.5 X TANDEM REPEATS.

```
FT REPEAT 283 322 1.
FT REPEAT 323 360 2.
FT REPEAT 361 398 3.
FT REPEAT 399 436 4.
FT REPEAT 437 451 5 (INCOMPLETE).
SQ SEQUENCE 657 AA; 71840 MW; 9B217001019C38A7 CRC64;

Alignment Scores:
Pred. No.: 5,4 Length: 657
Score: 95.00 Matches: 61
Percent Similarity: 35.77% Conservative: 27
Best Local Similarity: 24.80% Mismatches: 93
Query Match: 6.88% Indels: 65
DB: 1 Gaps: 12

US-09-658-824-808 (1-781) x MY16_MOUSE (1-657)
QY 55 GCTACTGACACAGCGGGGTAGTCCACAGGAGATCCAACTGGGAGTTGAAGTGTGAGT 114
Db 212 AlaThrGluGluGlyThrGluAsnLysAlaAspProSerAsnSerProSer 229
QY 115 GAGAGTGAAGAGAACACAGCAGCTCCGGAGGGTGTGTGTCAGTCACTCAGAGTGAG 174
Db 230 -----SerGlySerHisSerArgAlaTrpGluTyrTyrSerArgGlu 243
QY 175 AAGGCCCTCGAAGTGTGCTCCTCTCATGCGGTGCCAGCCCATGGACCTTCTTGTCTC 234
Db 244 LysProLysGlnGluGlyGluAlaLysValGluAlaHisArgAlaGlyGlnGlyHisPro 263
QY 235 GTCACGGCCATAACTAGGAGGAGGAGGAGGCGCGAGG-----AGTGGAGGG 279
Db 264 CysArgAsnAlaGluAlaGluGlyGlyProGluThrPheValCysThrGlyAsn 283
QY 280 GCTCAGGCGAAGTGGGTGCTGTGGGGGTATCCGAGTCCAGAGCAGCCTGGAAACCCC 339
Db 284 AlaPheLeuLys-----AlaTrpValTyr-Arg-----ProGlyGluAs 296
QY 340 GACAGCAATTCGAGTCC-----CCAGAGCGGACACAGGAGGAGGCGCATGAGCGA 393
Db 296 pThrGluGluGlu-AspAsnSerAspSerAspSerAlaGluGluAspThrAla 313
QY 394 CACACAAACACAGAACACACAGCCAGTCC-----ACACCTGGCGAGC 498
Db 314 --GlnThrGlyAlaThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProG 333
QY 426 -----CAGAGCCCAAGTATGAGAGCCCAAAAGAGAACACACAGCAGTG 471
Db 333 LysGluAspThrGluGluGluAspSerAspSerAlaGluGluAspThrAlaGlnT 353
QY 472 AAGTCGGGATCCCT-----ACACCTGGCGAGC 498
Db 353 hrGlyAlaThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyGlu 373
QY 499 AGACAGAGAGATCAGATACACTGATCCAGTCCGCGGACATGGAAGGTGATCTGC 558
Db 373 spThrGluGluGluAsnSer-----AspLeuAspSerAlaGlu 386
QY 559 AAGAGCTGCATCAGTCAAC-----ACCGGGGATAATCTGGATT-----GGGTTC 606
Db 386 LuAspThrAlaGlnThrGlyAlaThrProHisThrSerAlaPheLeuLysAlaTrpValT 406
QY 607 GGCCTCAAGTGAAGATAATACCTAAAGAGACACTGTAAATGCCAGAGCAGTGA 666
Db 406 YrArgProGlyGluAsp---Thr-GluGluGluAsnSerAspLeuAspSerAlaGluGlu 424
QY 667 GAGCAACACAA 678
Db 425 AspThrAlaGln 428

RESULT 21
ID YFHA_BORPE
AC P33445; STANDARD; PRT; 320 AA.
```

```
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 33.8 kDa protein in PHAC 3' region (ORFA).
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RA Willems R.J.L.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X64876; CAA46093.1; -
CC Hypothetical protein.
KW SEQUENCE 320 AA; 33781 MW; 1B5C54B11DBF2D73 CRC64;
SQ SEQUENCE 320 AA; 33781 MW; 1B5C54B11DBF2D73 CRC64;

Alignment Scores:
Pred. No.: 7,91 Length: 320
Score: 92.50 Matches: 45
Percent Similarity: 37.79% Conservative: 20
Best Local Similarity: 26.16% Mismatches: 65
Query Match: 6.70% Indels: 42
DB: 1 Gaps: 8

US-09-658-824-808 (1-781) x YFHA_BORPE (1-320)
QY 122 AAGAGAACACAGCAGCTCCGAGGTTGTGTGTGTGTCAG-----TGACTC 166
Db 142 ArgArgAsnGlnGlnAlaGlnGlyThrGlyAlaGlyGlnArgAlaGluAspLeuValLeu 161
QY 167 AGACTGAGAGGCCCTCGAAGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 226
Db 162 ArgIle-AlaAlaProArgGlnPheGlyAspGlyHis----- 173
QY 227 CTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 286
Db 174 -LeuAlaHisArgGlyAlaGly---GlyGlyArgArgAlaGlyAsn---ArgGlyGluAs 191
QY 287 CGAAGCTGGGGTGTGTGGGGGTATCCGAG---TCCCAGAGCAGCCTGGAACCCGACA 343
Db 191 pGlyAlaAlaAspValGlyValGlnGlnProAlaGlyGlnProValGlnProAlaGAr 211
QY 344 GAAGATTCTGGACTCCCGCAGCGGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 403
Db 211 gGluAlaLeuGluGlnValLeuGlyGlnAlaGlyAlaGluGlnAspLeuAlaHisProGl 231
QY 404 CACAGAACACACAGCAGCTCCAGGAGCCAGTAATGGAGAGCCCAAAAGAA----- 458
Db 231 nGluGlnArgGlnGlyGlnArgProAlaArgGlyAlaProAspArgAspGlyHis 251
QY 459 -----GAACACGACCTGAAAGTCCGGATCTCTACACCTGGCAG----- 497
Db 251 sGlyValAlaGlyArgProAlaGlyGluAspPheHisAlaAspProGlyArgAlaGlyGl 271
QY 498 -----CAGACAGAGAGATCAGATACAGTACAGTACAGTACAGTACAGTACAGTACAGT 529
Db 271 nGlyGluAlaAsnProAspSerAlaThrGlnGlnArgGluAspArgAsp----- 287
QY 530 CCCAGTCCCGCAGATCGAAGGTGATCTGCAAGAG 563
Db 288 -----AspGlnGlnGlyAspAspGlnAsp 295

RESULT 22
ID YFHA_BORPE
AC P33445; STANDARD; PRT; 320 AA.
```



Db 62 GlyAlaSerAlaGlyGlnGlyProLysPro---GluAlaHisSerGlnGluGlnGlyHis 80  
 QY 573 TCAACACCGGGGATAAATCTGGATTCTCCGGCTCAAGGTGAAGATAAATACCTAA 632  
 Db 81 ProGlnThrGlyCysGluCysGluAspGlyProAspGlyGlnGluMetAspPro-ProAs 100  
 QY 633 AGAGGAACACTGTAAATGCCAGAGAGAGGTGAAGAGCAACCAACAA 678  
 Db 100 nProGluGluValLysThrProGluGluGlyGlnLysSerGln 115

RESULT 24  
 TAU\_PAPHA  
 ID TAU\_PAPHA STANDARD; PRT; 382 AA.  
 AC Q9MYX8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Microtubule-associated protein tau (Neurofibrillary tangle protein)  
 DE (Paired helical filament-tau) (PHF-tau).  
 GN MAPT OR TAU.  
 OS Papio hamadryas (Hamadryas baboon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopitheciinae; Papio.  
 OX NCBI\_TaxID=9557;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Frontal cortex;  
 RA Wang X.L., Wang J., Schultz C., Hubbard G.B.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

-!- FUNCTION: PROMOTES MICROTUBULE ASSEMBLY AND STABILITY, AND MIGHT BE INVOLVED IN THE ESTABLISHMENT AND MAINTENANCE OF NEURONAL POLARITY. THE C-TERMINUS BINDS AXONAL MICROTUBULES WHILE THE N-TERMINUS BINDS NEURAL PLASMA MEMBRANE COMPONENTS, SUGGESTING THAT TAU FUNCTIONS AS A LINKER PROTEIN BETWEEN BOTH. AXONAL POLARITY IS DETERMINED BY TAU LOCALIZATION (IN THE NEURONAL CELL) IN THE DOMAIN OF THE CELL BODY DEFINED BY THE CENTROSOME (BY SIMILARITY).  
 -!- SUBCELLULAR LOCATION: MOSTLY FOUND IN THE AXONS OF NEURONS, IN THE CYTOSOL AND IN ASSOCIATION WITH PLASMA MEMBRANE COMPONENTS (BY SIMILARITY).  
 -!- TISSUE SPECIFICITY: EXPRESSED IN NEURONS.  
 -!- DOMAIN: THE TAU/MAP REPEAT BINDS TO TUBULIN.  
 -!- PTM: PHOSPHORYLATION AT VARIOUS SERINE AND THREONINE RESIDUES IN S-P OR T-P MOTIFS BY PROLINE-DIRECTED PROTEIN KINASES (PDPK: CDC2, CDK5, GSK3, MAPK) (A FEW SITES PER PROTEIN IN INTERPHASE, MORE IN MITOSIS), AND AT SERINE RESIDUES IN K-X-G-S MOTIFS BY MAP/MICROTUBULE AFFINITY-REGULATING KINASE (MARK) (BY SIMILARITY).  
 -!- SIMILARITY: CONTAINS 4 TAU/MAP REPEATS.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; AF281310; AAF97596.1;  
 InterPro; IPR002955; Tau\_protein.  
 DR InterPro; IPR001084; Tubulin-bind.  
 DR Pfam; PF00418; tubulin-binding; 4.  
 DR PRINTS; PR01261; TAUPROTEIN.  
 DR PROSITE; PS00229; TAU\_MAP; 4.  
 KW Microtubules; Cytoskeleton; Repeat; Acetylation; Phosphorylation.  
 FT INIT\_MET 0  
 FT REPEAT 185 215 TAU/MAP MOTIF 1.  
 FT REPEAT 216 246 TAU/MAP MOTIF 2.  
 FT REPEAT 247 277 TAU/MAP MOTIF 3.  
 FT REPEAT 278 309 TAU/MAP MOTIF 4.  
 FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
 FT DISULFID 232 263 BY SIMILARITY.  
 SQ SEQUENCE 382 AA; 39879 MW; D2D15A53AA00E87 CRC64;

Alignment Scores: 8.78 Length: 382  
 Pred. No.: 92.00 Matches: 59  
 Score: 40.38% Conservative: 25  
 Percent Similarity: 28.37% Mismatches: 79  
 Best Local Similarity: 6.67% Indels: 45  
 Query Match: 1 Caps: 11  
 DB: 1

US-09-658-824-808 (1-781) x TAU\_PAPHA (1-382)

QY 76 GGTCCACAGCGAGATCAACTGGGAGTTGAAGTGTGAGTGAAGAGGAACCAACGCA 135  
 Db 41 GlyLeuLysAlaGluGluAlaGlyIleGlyAspThrProSerLeuGluAspGluAlaAla 60  
 QY 136 GGTCTCCGAGGGTGTGTGGTCACTCAGTCACTGAGAGGAGCCCTCGAAGTCGTGTC 195  
 Db 61 Gly-----HisValThrGlnAla-ArgMetValSerLysSerLysAs 74  
 QY 196 CCTCTCATCGGTGCCACGCCCATGGACCTTCTGTCTCTCAGCGGCCATACTAGGAG 255  
 Db 74 p-----GlyThrGlySerAsp-----As 80  
 QY 256 GAAGAGGGGCGGAGGAGTGGAGGGCTCAGGCCGA----- 289  
 Db 80 pLysLysAlaLysGlyAlaAspGlyLysThrLysIleAlaThrProArgGlyAlaAlaPr 100  
 QY 290 -----AGTGGGTCTGTGGGGGTATCCGAGTC-----CCAGAAGCAAC 330  
 Db 100 oProGlyGlnLysGlyGlnAlaAsnAlaThrArgIleProAlaLysThrProAlaPr 120  
 QY 331 TGGAAACCCACACAGAAATCTGGACTCCACAGAGCGGACGAGAGGAGCGCATGAG 390  
 Db 120 oLysThrPro---ProSerGlyGluProLysSerGlyAspArgSerGlyTyrSe 139  
 QY 391 CGACACACACAAACACA---GAACACACACAGCCAGTCCAGAGCCAGTAATGGAGAGC 447  
 Db 139 r-SerProGlySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrP 159  
 QY 448 CCAAAAAGAAGAACCA-----GCAGCTCAAGTCCGGATCCTACACCTGGCGCAGAGA 501  
 Db 159 roProAlaArgGluProLysLysValAlaValArgThrProProLysSerProSerS 179  
 QY 502 CAGAAGAAGATCAG---GATACAGCTCAGATCCAGTCCGCGACATGGAAGGTGATCTGC 558  
 Db 179 erAlaLysSerArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal--- 197  
 QY 559 AAGAGCTGCATAGTCAACACCGGGGATAAATCTGGATTGGTTCGCGCGTCAAGGTG 618  
 Db 198 LysSerLysIleGlySerThrGluAsnLeuLysHisGlnProGlyGlyGly---LysVal 216  
 QY 619 AAGATAATACCTAAAGAG 636  
 Db 217 GlnIlelleAsnLysLys 222

RESULT 25  
 TISD\_HUMAN  
 ID TISD\_HUMAN STANDARD; PRT; 482 AA.  
 AC P47974;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Butyrate response factor 2 (TISILD protein) (EGF-response factor 2) (ERF-2).  
 DE BRF2 OR TISILD OR ERF2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96132724; PubMed=8545129;  
 RA Ino T., Iisui H., Hirano M., Kurosawa Y.;  
 RT \*Identification of a member of the Tis1l early response gene family

[illegible]







Db	649	LysGlyGlyPhe- PheSerSerPheMetLysLysArgAsnAlaPro----	ThrProProLys	667
QY	96	TGGAGCTTCAAGCTGTGACGTGACAGTGAAGAGCAACCA-----	-GC	134
Db	667	sArgSerSerPheArgGluMetGluAsnGlnProHisLysLysTyrrGluLeuThrG1	687	
QY	135	AGCCTTCGGAGGTTGTG-----	TGCTCAGTGA	173
Db	687	yAsnPheSerSerValAlaSerLeuGlnHisAlaAspGlyPheSerPheThrProAlaG1	707	
QY	174	GAAGGCCCTCGAAGTGTGTCCTCTCATCGGTGCCCGCCATCGACCTTCTGTCT	233	
Db	707	nGlnGluAlaAsnLeuValProPro-----LysCys-----	TyrGlyGlySerPheAl	723
QY	234	CGTCAGCGCCATACTAGGAGGAAGGAGGCCGAGAGTGGAGGGCTCAGGCGAAGCT	293	
Db	723	aglnArgAsnLeuCysAsnAspGlyGly-----	GlyGlyGlyGlySerGlyThrAl	741
QY	294	GGGGTCTCTGG-----GGGTAT-----	-----	312
Db	741	aglyGlyGlyTrpSerGlyIleThrGlyPhePheThrProArgLeuIleLysLysThrLe	761	
QY	313	-----	-----CCGAGTCCCAAG	326
Db	761	uGlyLeuArgAlaGlyLysProThrAlaSerAspAspThrSerLysProPheProArgSe	781	
QY	327	CACCTGGGAACCCGACAGACATTCTGCATCTCCCGACGGGACACAG-----	-----	373
Db	781	rasn-SerThrSerSerMetSerSerGlyLeuProGluGlnAspArgMetAlaMetThrL	801	
QY	374	-----	-----GAGAGGACGCATGACGCGACACAC	398
Db	801	euProArgAsnCysGlnArgSerLysLeuGlnLeuGluArg-Thr-----	-----ValSer	817
QY	399	ACAAACACAGAACACACACAGCCAGTCCAGGACCCAGTAAATGGAGACCCCAAAA	--- 455	
Db	818	ThrSerSerGlnProGluGluAsnValAspArgAlaAsnAspMetLeuProLysLysSer	837	
QY	456	GAAGAACACAGCAGCTGAAGCTGGGATCTTACACCTGGCGCAGACAGCAAGAACATCAG	515	
Db	838	GluGluSerAlaAlaProSerArgGluArgProLysAlaLysLeuLeu-----	-----ProArg	855
QY	516	GATACAGCTGAGATCCAGCTGCGGCACATGGAAGGTGATCTGCAAGAGTGTCATCAGTCA	575	
Db	856	GlyAlaThrAlaLeuProLeuArgThrProSerGlyAspLeuAlaIleThrGluLys-As	875	
QY	576	AACACCGGGATAAATCTGCATTTGGTTCCGGCGTCAAGGTGAAGATAAATACCTAAAGA	635	
Db	875	pProGlyValGlyVal-----	-----AlaGlyValAlaAlaProLysGlyLysG1	892
QY	636	GGAACAC-----	-----TGTAATAATGCCAAGCAGGTGAAGAG	669
Db	892	uLysAsnGlyGlyAlaArgLeuGlyMetAlaGlyValProGluAspGlyGluGln	910	
RESULT 29				
OTX1_HUMAN				
AC	ID	OTX1_HUMAN	STANDARD;	PRT; 354 AA.
AD	IC	P32242;		
DT	DT	01-OCT-1993 (Rel. 27, Created)		
DT	DT	01-OCT-1993 (Rel. 27, Last sequence update)		
DT	DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	GN	Homeobox protein OTX1.		
DE	GN	OTX1.		
OS	OS	Homo sapiens (Human).		
OC	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI_TaxID=9606;				
FN	FN	[1]		
RP	RP	SEQUENCE FROM N.A.		
RX	RX	MEDLINE-93327763; PubMed-8101484;		
RA	RA	Siameone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,		
RA	RA	Nigro V., Boncinelli E.;		
RT	RT	*A vertebrate gene related to orthodenticle contains a homeodomain of		

RT	the bicoid class and demarcates anterior neuroectoderm in the
RT	gastrulating mouse embryo.*;
RL	EMBO J. 12:2735-2747(1993).
CC	-!- FUNCTION: PROBABLY PLAY A ROLE IN THE DEVELOPMENT OF THE BRAIN AND
CC	THE SENSE ORGANS. CAN BIND TO THE BCD TARGET SEQUENCE (BTS):
CC	5'-TCTAATCC-3'.
CC	-!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC	-!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN.
CC	-!- DEVELOPMENTAL STAGE: EMBRYO.
CC	-!- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
CC	"BICOID" SUBFAMILY.
DR	HSSP: P22808; 1VND.
DR	TRANSFAC: T02079; -.
DR	MIM: 600036; -.
DR	InterPro: IPR001356; Homeobox.
DR	InterPro: IPR003025; Otx_TF.
DR	Pfam: PF00046; homeobox; 1.
DR	PRINTS: PR00024; HOMEBOX.
DR	PRINTS: PR01255; OTXHOMEBOX.
DR	SMART: SM00389; HOX; 1.
DR	PROSITE: PS00027; HOMEBOX_1; 1.
DR	PROSITE: PS00071; HOMEBOX_2; 1.
KW	Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT	DNA_BIND 38 97 HOMEBOX.
FT	DOMAIN 275 301 HIS-RICH.
FT	SEQUENCE 354 AA; 37327 MW; E32C1E4746BDAA CRC64;
SQ	
Alignment Scores:	
Pred. No.:	9.53
Score:	Length: 354
Percent Similarity:	Matches: 50
Best Local Similarity:	Conservative: 20
Query Match:	Mismatches: 59
DB:	Indels: 63
	Gaps: 10
US-09-658-824-808 (1-781) x OTX1_HUMAN (1-354)	
Qy	622 TCTTCACCTGAGCGCGGAACCAATCCAGATTATCCCGGTGTTGACTGATGCAGC 563
Db	112 SerSerProValargGlusSerGlySerGluSerSerGlyGlnPhe----- 127
Qy	562 TCTTGCAGATCACCTTCCCATGTGCGGCACCTGGGATCTCAGCTGATCTGATCTCTTCT 503
Db	128 -----ThrProAlaValSerSerAlaSerSerSerSerSerSerSerSerSerSer 145
Qy	502 GTCTGCTGCCAGGTAGATCCCGACTTTCAGCTGCTGTTCTCTTTTGGGCTCT 443
Db	146 SerAlaasnProAla-----AlaAlaAlaAlaAlaGlyLeuGlyGlyAsn 160
Qy	442 CCATTACTGGGCTCTCTGGGACTGGCTGTGTGTTGTTGTCGCTCATGCC 383
Db	161 Pro-----ValAlaAlaAla 165
Qy	382 GTCCCTCTCCTGGTCCGCTCTGGGAGTCCAGAACTTCTGTGCGGGTTCCAGGTGCTTC 323
Db	166 SerSerLeuSerThrProAlaAlaSerSer-----Ile 176
Qy	322 TGGGAC-----TCGGATACCCCAACACACCCAGCTTCG-----CCTGAGGCC 278
Db	177 TrpSerProAlaSerIleSerProGlySerAlaProAlaSerValSerValProGluPro 196
Qy	277 CTC-----CACTCTCGGCCCTCTCTCCTCCCTAGTTATGCGCGTGACGACAGAC 230
Db	197 LeuAlaAlaProSerAsnThrSerCysMet-----GlnArg 208
Qy	229 AAGAAGGTCATGGCGGTGGCAGCCGATGAGAGGGACGACGACTT-----CGAGGG 179
Db	209 SerValAlaAlaGlyAlaAlaThrAlaAlaAlaAlaSerTyrProMetSerTyrGlyGlnGly 228
Qy	178 CCTTCTCACTCTGAGTCACTGACCAACACACCCCTCC----- 143
Db	229 GlySerTyrGlyGlnGlyTyrProThrProSerSerSerTyrPheGlyValAspCys 248

[illegible]

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AB013729; BAA76294.1; -  
 CC MGD: MGI:1338032; Sema6c.  
 CC InterPro: IPR001627; Sema.  
 CC Pfam: PF01403; Sema; 1.  
 CC Signal; Transmembrane; Multigene family; Neurogenesis; Glycoprotein;  
 KW Developmental protein.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 931 SEMAPHORIN 6C.  
 FT DOMAIN 26 605 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 606 626 POTENTIAL.  
 FT DOMAIN 627 931 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 234 541 SEMA.  
 FT DOMAIN 663 669 POLY-PRO.  
 FT DOMAIN 754 757 POLY-PRO.  
 FT CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 438 438 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 931 AA; 99537 MW; B0D99D594209F125 CRC64;

## Alignment Scores:

Pred. No.: 14.8 Length: 931  
 Score: 89.50 Matches: 60  
 Percent Similarity: 32.23% Conservative: 18  
 Best Local Similarity: 24.79% Mismatches: 70  
 Query Match: 6.28% Indels: 95  
 DB: 1 Gaps: 13

US-09-658-824-808 (1-781) x SM6C\_MOUSE (1-931)

Qy 638 TCCTCTTTAGTATATCTTACCTTCAGCGCGGACCCAAATCCAGATTATCCCGGT 579  
 Db 595 SerAlaThrGlyValArgAspLeuSerProAlaSerProAlaSerArgSerIlePro--- 603  
 Qy 578 GTTTGACTGATGACGCTCTTGACATCACCTTCATGCGGCACTGGGATCAGCTGT 519  
 Db 604 -----1leProLeuLeu-----LeuAlaCys 610  
 Qy 518 ATCTGATCTCTCTCTGCTGCCAGGTAGGATCCCGACTTTCAGCTGCTGTTTC 459  
 Db 611 ValAlaAlaAlaPheAla----- 616  
 Qy 458 TTTCTTTTGGGGCTCTCCATTAATGCTGCTGGGACTGCTGTGTTCTGTTTGT 399  
 Db 617 -----LeuGlyAlaSerValSerGlyLeuLeu-----ValSerCys 628  
 Qy 398 GTGTGCTGCTCAGC-----CCT 381  
 Db 629 AlaCysArgArgAlaAsnArgArgSerLysAspIleGluThrProGlyLeuProArg 648  
 Qy 380 CCCTC-----TCCTGGTCCCTCTG-----GGAGTCCAGAA----- 349  
 Db 649 ProLeuSerLeuArgSerLeuAlaArgLeuHisGlyGlyProGluProProProPro 668  
 Qy 348 -----TCCTCTGCTGGGGTTCAGGTGCTTCTGGGACTCGGATACCCCA--- 304  
 Db 669 ProLysAspGlyAspAlaAlaGlnThrProGlnLeuTyThrThrPheLeuProProPro 688  
 Qy 303 -----ACAGCACCCAGCTCGCTGACGCCCTCCAGCTCTCGGCTCTCTCTCCTC 253  
 Db 689 AspGlyGlySerProProGluLeuAlaCysLeuProThrProGluThrThrProGluLeu 708

Qy 252 CCTAGTTATGGCCGTGACGAGACAAGAGGTCCATGGCGTGCCAGCCGATCAGAGGAC 193  
 Db 709 Pro-VallYsHisLeuArgAlaSerGlyGlyProThrGluTrpAsnGln---AsnGlyAs 727  
 Qy 192 GACGACTTCGAGGGCTCTTCTACTCTGAGTCTACCAACCCCTCCGGAACCTGTC 133  
 Db 727 nAsnAlaSerGluGlyProGly-----ArgProProAr 738  
 Qy 132 TGGTTCCTCTTCTACTCTCTACTCTACACTTCAACTCCAGTTCGCTGCGACCTAC 73  
 Db 738 gGlyCysSer-----GlyAlaGlyGlyProAl 747  
 Qy 72 CCGCGTGTCTCAGTAGCGGAGAGAAGATCCAGACTTCAGGACCGAGTCCGCGGTCA 13  
 Db 747 aproArgValLeuVal-----ArgProProProGlyCysProGlyG 762  
 Qy 12 CAGC 9  
 Db 762 nAla 763  
 RESULT 32  
 MAN1\_MOUSE STANDARD; PRT; 331 AA.  
 ID AC Q9WU40;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Inner nuclear membrane protein Man1 (Fragment).  
 GN MAN1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20138223; PubMed=10671519;  
 RA Lin F., Blake D.L., Callebaut I., Skerjanc I.S., Holmer L.,  
 RA McBurney M.W., Paulin-Levasseur M., Worman H.J.;  
 RT "MAN1, an inner nuclear membrane protein that shares the LEM domain  
 RL J. Biol. Chem. 275:4840-4847(2000).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER NUCLEAR  
 CC MEMBRANE.  
 CC -1- SIMILARITY: CONTAINS 1 LEM DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF112300; AAD31594.1; -  
 CC InterPro: IPR003887; LEM.  
 CC Pfam: PF03020; LEM; 1.  
 DR SMART; SM00540; LEM; 1.  
 KW Nuclear protein; Transmembrane.  
 FT DOMAIN 8 51 LEM.  
 FT DOMAIN 52 62 POLY-GLN.  
 FT DOMAIN 74 79 POLY-ASN.  
 FT DOMAIN 151 157 POLY-GLY.  
 FT DOMAIN 264 272 POLY-GLY.  
 FT NON\_TER 331 331  
 SQ SEQUENCE 331 AA; 34594 MW; BB7AD1BD85FCF54D CRC64;  
 Alignment Scores:  
 Pred. No.: 14.8 Length: 331  
 Score: 89.00 Matches: 51  
 Percent Similarity: 39.23% Conservative: 20  
 Best Local Similarity: 28.18% Mismatches: 62  
 Query Match: 6.45% Indels: 48  
 DB: 1 Gaps: 9

```
US-09-658-824-808 (1-781) x MAN1_MOUSE (1-331)
QY 31 CCCTCAGGCTCGGATCTTCTCCCTACTGAGACACGGCGGTAGTCCACAGGCAGAT 90
    ||| ||||| |||||: ||| |||
Db 99 ProAlaGlyLeuGlyArgLeuSerAlaSerAlaAlaGlu-----SerProValAlaGly 116
QY 91 CCNACTGGAGTTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 150
    :||| |||||: ||||| |||||
Db 117 GlySerGlyGly-----AlaAlaAlaValProAlaAlaGlySerLysVal 131
QY 151 GTGTGCTCAGTGTGACTCAGAGTGTGAGAGGCCCTCGAAGTGTGCTGCTCTCATCGGTGC 210
    :||| ||||| |||||: |||||
Db 132 LeuLeuGlyPheSerSerAspGluSerAspValGlu----- 143
QY 211 CAGGCCCATGAGCTTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 267
    ||| |||||: ||||| |||||
Db 144 -----AlaSerProArgGluGluAlaGlyGlyGly 153
QY 268 AGGAGTGTGAGGGCTCAGGCGAAGTGTGGTGTCTGT-----GGGGTATCCGAGTCC 321
    ||| ||||| |||||: ||||| |||||
Db 154 GlyGlyGlyGlyAla-ArgArgAspArgAlaAlaLeuGlnTyrArgGlyLeuArgAlaPr 173
QY 322 AGAAGCAGCTGGAACCCCGACAGAGATTTCTGGACTCCCGACAGCGGACAGGAGGGA 381
    ||| ||||| |||||: ||||| |||||
Db 173 opRoAlaProAlaAlaGlyGluValThrGly-GlyHisProGlyGluArgArg----- 191
QY 382 CGCGATGAGCGACACACACACACAGAGACACACAGC-----CAGTCCAGGAGGCCA 435
    :||| ||||| |||||: ||||| |||||
Db 192 -----LysProHisSerTrpTrpGlyAlaArgArgProA 203
QY 436 GTAATGGAGAGCCCCAAAAGAACAGCAGCTGAAAGTCGGGATCCTACAGCTGGCC 495
    :||| ||||| |||||: ||||| |||||
Db 203 laGlyProGluProGln-----ProProAlaAlaGlySer-----AspGlyA 217
QY 496 AGCAGACAGACAGATCAGATCAGCTGAGATCCCGAGTCCCGAGCATGGAAGGTGT 554
    :||| ||||| |||||: ||||| |||||
Db 217 laAlaGluAspAlaAspGluLeuAlaAspGlyGluAspArgAspProGluAlaGlu 236

RESULT 33
ID S24B_ARATH STANDARD; PRT: 1069 AA.
AC Q9M081; O65535;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative protein transport protein Sec24-like At4g32640.
GN AT4G32640 or F4p11.160.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=20083486; PubMed=10617199;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
RA Harris B., Ansong W., Brand P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Bernelstein S., Hempel S., Feldpausch M., Lambert S., Van den Baele H.,
RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
```

```
Clark L., Doggett J., Hall S., Key M., Lennard N., McIlroy K., Mayes R.,
Pettitt A., Rajandream M.A., Lyne M., Benes V., Rehnemann S.,
Borkova D., Blocker H., Scharfe M., Grimm M., Loehner T.-H.,
Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
Neumann S., Argirou A., Vitale D., Liquori R., Piravandi E.,
Massenet O., Hugier F., Clabault G., Muendlein A., Felber R.,
Schmahl S., Hillier R., Schmidt W., Lecharny A., Aubourg S.,
Cheifod F., Cooke R., Berger C., Monfort A., Casacuberta E.,
Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Jesse T.,
Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Bleke C.,
Heijnen L., Schwarz S., Scholler P., Heber S., Frances P., Bliecke C.,
Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
Zaccaria P., Bedan M., Wilson R.K., de la Bastide M., Habermann K.,
Parnell L., Devina N., Gnoj L., Schutz K., Huang E., Spiegel J.,
Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
Stonking T., Kalicki J., Graves T., Harmon G., Edwards J.,
Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
Antonoli B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
Granat S., Shohdy N., Hasegawa A., Rameed A., Lodhi M., Johnson A.,
Chen E., Marra M., Martienssen R., McCombie W.R.;
*Sequence and analysis of chromosome 4 of the plant Arabidopsis
Nature 402:769-777(1999).
-!- FUNCTION: COMPONENT OF THE COPII COAT, THAT COVERS ER-DERIVED
VESICLES INVOLVED IN TRANSPORT FROM THE ENDOPLASMIC RETICULUM TO
THE GOLGI APPARATUS. COPII IS COMPOSED OF AT LEAST FIVE PROTEINS:
THE SEC23/24 COMPLEX, THE SEC13/31 COMPLEX, AND THE PROTEIN SAR1.
ACTS IN THE CYTOPLASM TO PROMOTE THE TRANSPORT OF SECRETORY,
PLASMA MEMBRANE, AND VACUOLAR PROTEINS FROM THE ENDOPLASMIC
RETICULUM TO THE GOLGI COMPLEX (BY SIMILARITY).
-!- SIMILARITY: BELONGS TO THE SEC23/SEC24 FAMILY. SEC24 SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: AL022537; CAAL8597.1; ALT_SEQ.
DR EMBL; AL161581; CAB79981.1; -.
KW Hypothetical protein; Transport; Protein transport; Golgi stack;
FT DOMAIN 430 455 ZINC FINGER-LIKE.
FT DOMAIN 55 60 POLY-PRO.
FT DOMAIN 341 344 POLY-SER.
FT DOMAIN 358 361 POLY-PRO.
SQ SEQUENCE 1069 AA; 114775 MW; 81CD4290B9FA3FD7 CRC64;

Alignment Scores:
Pred. No.: 16.4 Length: 1069
Score: 89.00 Matches: 58
Percent Similarity: 34.72% Conservative: 17
Best Local Similarity: 26.85% Mismatches: 52
Query Match: 6.24% Indels: 89
DB: 1 Gaps: 13

US-09-658-824-808 (1-781) x S24B_ARATH (1-1069)
QY 643 AGTCTCTCTCTTAGGTATTATCTTCACCTTCACCCGCGAACCACCAAT----- 596
    :||| ||||| |||||: ||| ||| |||||
Db 112 AsnValProLeuAsnArgProThrGlyProSerArgGlnProSerPheGlySerArg 131
QY 595 CCAGATTATA-----TCCCGCGTGTGTTGACTGATGTCAGCTCTTCGATGATGATG 542
    ||| :||| ||||| ||||| ||||| ||||| |||||
```









```
QY 290 -----AGCTGGGGTCTGTTGGGGGTATCCGAGTCCCA----- 322
Db 158 oProGlycInLysGlyGlnAlaAsnAlaThrArgIleProAlaLysThrProProAlaPr 178
QY 323 -----GAAGCACTGGAAACCC 339
Db 178 oLysThrProProSerSerAlaThrLysGlnValGlnArgLysProProProAlaGluPr 198
QY 340 GACAGAGATTCGGACTCCCGACCGGACGAGGAGGAGCGGATGAGCGACACACA 399
Db 198 oThrSerGluArgGlyGluProProLysSerGlyAspArgSerGlyTrSer-SerProG 218
QY 400 CAACACAC---GAACACACAGCAGTCCCGAGGAGCCAGTAATGGAGAGCCCAAAAG 456
Db 218 lySerProGlyThrProGlySerArgSerArgThrProSerLeuProThrProProAla 238
QY 457 AAGAACCA-----GCAGCTGAAGTGGGATCCTACACCTGGCGACGACAGAGAAG 510
Db 238 rgGluProLysLysValAlaValAlaArgThrProProLysSerProSerSerAlaLys 258
QY 511 ATCAG---GATACAGCTGAGATCCCGAGCGGACATGGAAAGTGATCTGCAAGAGCTGC 567
Db 258 erArgLeuGlnThrAlaProValProMetProAspLeu-LysAsnVal---LysSerLys 276
QY 568 ATCAGTCAAAACCGGGGATTAATCTGGATTGGGTTCCGGCTCAAGGTGAAGATAATA 627
Db 277 IleGlySerThrGluAsnLeuLysHisGlnProGlyGlyGly---LysValGlnIle 295
QY 628 CCTAAAGAG 636
Db 296 AsnLysLys 298

RESULT 38
G19P_BOVIN
ID G19P_BOVIN STANDARD; PRT; 533 AA.
AC Q28034;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Protein kinase C substrate, 60.1 kDa protein, heavy chain (PKCSH)
DE (80K-H protein).
GN PKCSH.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HOLSTEIN; TISSUE=Corpus luteum;
RA Brule S., Rabahi F., Beckers J., Monniaux D., Silversides D.W.,
RA Lussier J.G.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U49178; AAA92060.1; -.
DR HSSP; P02593; IAK8.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000886; ER_target.
DR InterPro; IPR002172; LDL_recept_A.
DR SMART; SM00192; LDLA; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR KW Phosphorylation.
FT DOMAIN 313 334 ASP/GLU-RICH (ACIDIC)..
SEQUENCE 533 AA; 60151 MW; 507858677PB81E01 CRC64;
```

```
Alignment Scores: 18.4 Length: 533
Pred. No.: 88.00 Matches: 46
Score: 35.75% Conservative: 28
Percent Similarity: 22.22% Mismatches: 77
Best Local Similarity: 6.38% Indels: 56
Query Match: 1
DB: 7

US-09-658-824-808 (1-781) x G19P_BOVIN (1-533)
QY 79 CCACGACGACATCAACTGGGAGTTGAAGTGTGAGTCAGAGTGAAGAG----- 126
Db 237 ProGluLeuAspThrAspGlyAspGlyAlaLeuSerGlyGluAlaGlnThrLeuLeu 256
QY 127 -----GAACACACAGCGCTTCGAGGAGGTTGTGTGTCAGTCAGTCAGTCAG 168
Db 257 GlyGlyAspAlaGlnMetAspAlaAlaPhePheTyArgValTrpAlaAlaIleArg 276
QY 169 AGTGAGNAGCCCTCGAAGTCGTCGTCCTCTCATCGGTGCCACGCCCATGGACCTCT 228
Db 277 AspLysTyArgSerGluValLeuProThr-----GluTyProProSer 291
QY 229 TGTCTCGTCACGGCCATACTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 288
Db 292 ProProAlaProAspVal-HetGluProLysGluGluGln----- 304
QY 289 AAGCTGGGGTGTGTTGGGGGTATCCGAGTCCACAGACACCTGGAACCCCGACAGAAGA 348
Db 305 -----ProProMetProSerProProThrGluGlu 314
QY 349 TTCTGGACTCCCGACAGCGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 408
Db 314 uGluAspGluAspGluGluGluGluGluGlu-----GluAspGluAspGluGlu-A 332
QY 409 AACCCACACAGCCAGTCCCGAGGAGCCAGTAATGGA-----GAGCCCAAA 453
Db 332 spGluAspSerGlnGlyGluGlnProLysAspAlaProProProAlaProAlaProGlnT 352
QY 454 AAGAAGAACCCAGCGTGAAGTCCGGATCCTACACTGGGCGACAGACAGAA----- 506
Db 352 hrAlaSerProThrGluGluAspArgMetProProTyArgGluGlnThrGlnAlaPheI 372
QY 507 -----GAAGATCAGGATACAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 543
Db 372 leAsnAlaAlaGlnGluAlaArgAsnLysPheGluGluAlaGluArgSerLeuLysAspM 392
QY 544 TGAAGGTGATCTCCAGAGAGTCATCATGCAACACACCGGGGATAAATCTGGATTGGGT 603
Db 392 etGluGluSerIleArgAsnLeuGluGln-----GluIleSerPheAsp 407
QY 604 TCCGGCGTCAAGGTGAA 620
Db 407 hegLyProAsnGlyGlu 412

RESULT 39
CIKE.DROME
ID CIKE.DROME STANDARD; PRT; 1174 AA.
AC Q02280;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Potassium channel protein eag.
GN EAG.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91262635; PubMed=1840699;
RA Warmke J., Drysdale R., Ganetzky B.;
```







PF 30-JUN-2000; 2000WO-US18061.  
XX  
XX  
XX 30-JUN-1999; 99US-0346492.  
PR 15-OCT-1999; 99US-0419356.  
PR 17-DEC-1999; 99US-0466867.  
PR 30-DEC-1999; 99US-0476300.  
PR 06-MAR-2000; 2000US-0519642.  
PR 22-MAR-2000; 2000US-0533077.  
PR 10-APR-2000; 2000US-0546259.  
PR 27-APR-2000; 2000US-0560406.  
PR 05-JUN-2000; 2000US-0589184.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;  
PI Retter MW, Mannion J;  
PI WPI; 2001-071488/08.  
XX  
XX Lung tumor-associated proteins and the nucleic acids that encode them,  
PT useful for preventing, diagnosing and treating lung cancer -  
PT  
XX  
XX Example 1; Page 427; 436pp; English.  
XX  
XX The present invention describes immunogenic portions of lung tumour-  
CC associated proteins (I) and the nucleic acids (NAs) that encode them.  
CC (I) have cytostatic activity and can be used in gene therapy, antisense  
CC inhibition and in vaccines. The NAs and the lung tumour-associated  
CC proteins they encode may be used in the prevention, treatment and  
CC diagnosis of diseases associated with their inappropriate expression,  
CC especially lung cancers. For example, the NAs may be administered to  
CC treat diseases by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of the protein by expressing inactive proteins  
CC or to supplement the patient's own production of (I). Additionally, the  
CC NAs may be used to produce the lung tumour associated protein, according  
CC to standard recombinant DNA methodology. Conversely, antisense NA  
CC molecules may be administered to down regulate protein expression by  
CC binding with the cells own genes and preventing their expression. The NA  
CC and complementary sequences may also be used as DNA probes in diagnostic  
CC assays to detect and quantitate the presence of similar NA sequences in  
CC samples, and hence which patients may be in need of treatment for lung  
CC cancer. The (I) may be used as antigens in the production of antibodies  
CC and in assays to identify modulators (agonists and antagonists) of the  
CC expression and activity of the protein. AAF68083 to AAF68878 and  
CC AAB76848 to AAB76878 represent human lung tumour protein related  
CC nucleotide and protein sequences which are used in the exemplification  
CC of the present invention.  
XX  
XX Sequence 160 AA:  
SQ  
Alignment Scores:  
Pred. No.: 6,7e-81 Length: 160  
Score: 887.00 Matches: 160  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 64.28% Indels: 0  
DB: 22 Gaps: 0  
US-09-658-824-808 (1-781) x AAB76877 (1-160)  
QY 202 ATGCGGTGCCAGCCGACCTTCTTGTCTCGTCACGGCATAACTAGGAGGAGGA 261  
Db 1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaIleThrArgGluGluGly 20  
QY 262 GGGCCGAGAGTGGAGGGCTCAGCGAGCTGGGGTGTGTGGGGGTATCCGAGTCCC 321  
Db 21 GlyProArgSerGlyGlyAlaGlnAlaLysLeuGlyCysCysTrpGlyTyrProSerPro 40  
QY 322 AGAAGCAGCTGAACCCGACAGAGATTCTTGGACTCCCGACGGGACGAGGAGGGA 381  
Db 41 ArgSerThrTrpAsnProAspArgPheTrpThrProGlnThrGlyProGlyGluGly 60  
QY 382 CGGCATGACGACACACAAACACAGAACACAGCCAGTCCCGAGGAGGCCAGTAATG 441

Db 61 ArgHisGluArgHisThrGlnThrGlnAsnHisThrAlaSerProArgSerProValMet 80  
QY 442 GAGAGCCCCAAAAGAAAGAACCCAGCAGCTGAAAGTCCGGATCCCTACACCTGGGCGACGAGA 501  
Db 81 GluSerProLysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 100  
QY 502 CAGAAGATCAGGATACAGCTGAGATCCAGTCGCGACATGGAAGGTGATCTGCAAG 561  
Db 101 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysValIleCysLys 120  
QY 562 AGCTGCATCAGTCAACACACCGGGGATAAATCTGCATTGGGTTCCGGCGTCAAGGTGAAG 621  
Db 121 SerCysIleSerGlnThrProGlyIleAsnLeuAspLeuGlySerGlyValLysValLys 140  
QY 622 ATAATACCTAAAGAGGACACTGTAAATGCCAGAACGAGGTGAAGAGCAACACCAAGTT 681  
Db 141 IleIleProLysGluGluHisCysLysMetProGluAlaGlyGluGluGlnProGlnVal 160  
RESULT 2  
AAB76869  
ID AAB76869 standard; Protein: 126 AA.  
XX  
XX AAB76869;  
XX  
XX 12-APR-2001 (first entry)  
XX  
XX Human lung tumour protein related protein sequence SEQ ID NO:791.  
DE  
XX Human; lung cancer; lung tumour; lung tumour protein; gene therapy;  
KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;  
KW cytostatic; antisense inhibition.  
XX  
XX Homo sapiens.  
XX  
XX WO200100828-A2.  
XX  
XX 04-JAN-2001.  
XX  
XX 30-JUN-2000; 2000WO-US18061.  
XX  
XX 30-JUN-1999; 99US-0346492.  
PR 15-OCT-1999; 99US-0419356.  
PR 17-DEC-1999; 99US-0466867.  
PR 30-DEC-1999; 99US-0476300.  
PR 06-MAR-2000; 2000US-0519642.  
PR 22-MAR-2000; 2000US-0533077.  
PR 10-APR-2000; 2000US-0546259.  
PR 27-APR-2000; 2000US-0560406.  
PR 05-JUN-2000; 2000US-0589184.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;  
PI Retter MW, Mannion J;  
PI WPI; 2001-071488/08.  
XX  
XX Lung tumor-associated proteins and the nucleic acids that encode them,  
PT useful for preventing, diagnosing and treating lung cancer -  
PT  
XX  
XX Example 1; Page 413; 436pp; English.  
XX  
XX The present invention describes immunogenic portions of lung tumour-  
CC associated proteins (I) and the nucleic acids (NAs) that encode them.  
CC (I) have cytostatic activity and can be used in gene therapy, antisense  
CC inhibition and in vaccines. The NAs and the lung tumour-associated  
CC proteins they encode may be used in the prevention, treatment and  
CC diagnosis of diseases associated with their inappropriate expression,  
CC especially lung cancers. For example, the NAs may be administered to  
CC treat diseases by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of the protein by expressing inactive proteins  
CC or to supplement the patient's own production of (I). Additionally, the  
CC NAs may be used to produce the lung tumour associated protein, according  
CC to standard recombinant DNA methodology. Conversely, antisense NA  
CC molecules may be administered to down regulate protein expression by  
CC binding with the cells own genes and preventing their expression. The NA  
CC and complementary sequences may also be used as DNA probes in diagnostic  
CC assays to detect and quantitate the presence of similar NA sequences in  
CC samples, and hence which patients may be in need of treatment for lung  
CC cancer. The (I) may be used as antigens in the production of antibodies  
CC and in assays to identify modulators (agonists and antagonists) of the  
CC expression and activity of the protein. AAF68083 to AAF68878 and  
CC AAB76848 to AAB76878 represent human lung tumour protein related  
CC nucleotide and protein sequences which are used in the exemplification  
CC of the present invention.

12

12



Score: 423.50		Matches: 101
Percent Similarity: 46.76%		Conservative: 0
Best Local Similarity: 46.76%		Mismatches: 2
Query Match: 30.69%		Indels: 114
DB: 22		Gaps: 2
US-09-658-824-808 (1-781) x AAB76870 (1-108)		
QY	2	CGCGGAGCTGTAGCCGCGACTCGGGTCCCTGAGGTCTGGATCTTTCTCCGCTACTG 61
DB	1	ArgArgSerCysGluProAlaThrArgValProGluValTrpLeuSerProLeuLeu 20
QY	62	AGACACGGCGGGTAGTCCACAGGCACATCCAACTGGGAGTTGAAGTGTGAGTGAGAGTG 121
DB	21	ArgHisGly----- 23
QY	122	AAGAGAACACAGCAGGCTTCGGAGGGTGTGTGTCAGTCAGTCAGAGTGAGAAGGCC 181
DB	23	----- 23
QY	182	TCGAAGTCGTCGCTCCCTCTCATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGG 241
DB	23	----- 23
QY	242	CCATAACTAGGAGGAGGAGGCGCGAGGAGTGGAGGGCTCAGCGGAGCTGGGGTGCT 301
DB	23	----- 23
QY	302	GTGTGGGGTATCCGAGTCCCAAGACACCTGGAAACCCCGACAGAAAGATTCTGGACTCCCC 361
DB	23	----- 23
QY	362	AGACGGGACAGGAGGAGCGGATGAGCGACACACAAACACACAGACACAGCCCA 421
DB	24	-----Gly--HisThrGlnThrGlnAsnHisThrAla 34
QY	422	GTCCAGGAGCCAGTAATGAGAGCCCCAAAAGAAAGAACACAGCAGCTGAAAGTCGGGA 481
DB	34	erProArgSerProValMetGluSerProLysLysLysAsnGlnLeuLysValGlyI 54
QY	482	TCCTACACTGGGACGACAGAGAGATCAGGATACAGCTCAGATCCAGTG----- 536
DB	54	leLeuHisLeuGlySerArgGlnLysLysIleArgIleGlnLeuArgSerGlnValLeuG 74
QY	537	-----CGCACATGAAGTGTATCTGCAAGAGCTGCATCAGTCAAAACACCGGG 585
DB	74	lyArgGluMetArgAspMetGluGlyAspLeuGlnGluLeuHisGlnSerAsnThrGlyA 94
QY	586	ATAAATCTGGATTTGGTTCGCGGCTCAAGGTGAAGATAATACC 629
DB	94	sPlysSerGlyPheGlyPheArgArgGlnGlyGluAspAsnThr 108
RESULT 4		
AAB76870		
ID	AAB76870 standard; Protein: 108 AA.	
XX	AAB76870;	
AC		
DT	12-APR-2001 (first entry)	
DE	Human lung tumour protein related protein sequence SEQ ID NO:793.	
XX		
KW	Human; lung cancer; lung tumour; lung tumour protein; gene therapy;	
KW	lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;	
KW	cytostatic; antisense inhibition.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200100828-A2.	
XX		
PD	04-JAN-2001.	
XX		
PF	30-JUN-2000; 2000WO-US18061.	

XX

PR 30-JUN-1999; 99US-0346492.

PR 15-OCT-1999; 99US-0419356.

PR 17-DEC-1999; 99US-0468667.

PR 30-DEC-1999; 99US-0476300.

PR 06-MAR-2000; 2000US-0519642.

PR 22-MAR-2000; 2000US-0533077.

PR 10-APR-2000; 2000US-0546259.

PR 27-APR-2000; 2000US-0560406.

PR 05-JUN-2000; 2000US-0589184.

XX (CORI-) CORIXA CORP.

XX PA Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;

XX PI Retter MW, Mannion J;

XX PT WPI: 2001-071488/08.

XX DR Lung tumor-associated proteins and the nucleic acids that encode them,

XX PT useful for preventing, diagnosing and treating lung cancer -

XX PS Example 1; Page 413-414; 436pp; English.

XX CC The present invention describes immunogenic portions of lung tumour-

XX CC associated proteins (I) and the nucleic acids (NAs) that encode them.

XX CC (I) have cytostatic activity and can be used in gene therapy, antisense

XX CC inhibition and in vaccines. The NAs and the lung tumour-associated

XX CC proteins they encode may be used in the prevention, treatment and

XX CC diagnosis of diseases associated with their inappropriate expression,

XX CC especially lung cancers. For example, the NAs may be administered to

XX CC treat diseases by rectifying mutations or deletions in a patient's genome

XX CC that affect the activity of the protein by expressing inactive proteins

XX CC or to supplement the patient's own production of (I). Additionally, the

XX CC NAs may be used to produce the lung-tumour associated protein, according

XX CC to standard recombinant DNA methodology. Conversely, antisense NA

XX CC molecules may be administered to down regulate protein expression by

XX CC binding with the cells own genes and preventing their expression. The NA

XX CC and complementary sequences may also be used as DNA probes in diagnostic

XX CC assays to detect and quantitate the presence of similar NA sequences in

XX CC samples, and hence which patients may be in need of treatment for lung

XX CC cancer. The (I) may be used as antigens in the production of antibodies

XX CC and in assays to identify modulators (agonists and antagonists) of the

XX CC expression and activity of the protein. AAF68083 to AAF68878 and

XX CC AAB76848 to AAB76878 represent human lung tumour protein related

XX CC nucleotide and protein sequences which are used in the exemplification

XX CC of the present invention.

XX SQ Sequence 108 AA;

Alignment Scores:

Pred. No.: 3,56e-34 Length: 108

Score: 423.50 Matches: 101

Percent Similarity: 46.76% Conservative: 0

Best Local Similarity: 46.76% Mismatches: 2

Query Match: 30.69% Indels: 114

DB: 22 Gaps: 2

US-09-658-824-808 (1-781) x AAB76870 (1-108)

QY 2 CGCGGAGCTGTAGCCGCGACTCGGGTCCCTGAGGTCTGGATCTTTCTCCGCTACTG 61

DB 1 ArgArgSerCysGluProAlaThrArgValProGluValTrpLeuSerProLeuLeu 20

QY 62 AGACACGGCGGGTAGTCCACAGGCACATCCAACTGGGAGTTGAAGTGTGAGTGAGAGTG 121

DB 21 ArgHisGly----- 23

QY 122 AAGAGAACACAGCAGGCTTCGGAGGGTGTGTGTCAGTCAGTCAGAGTGAGAAGGCC 181

DB 23 ----- 23

QY 182 TCGAAGTCGTCGCTCCCTCTCATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGG 241

QY 242 CCATAACTAGGAGGAGGAGGCGCGAGGAGTGGAGGGCTCAGCGGAGCTGGGGTGCT 301

QY 302 GTGTGGGGTATCCGAGTCCCAAGACACCTGGAAACCCCGACAGAAAGATTCTGGACTCCCC 361

QY 362 AGACGGGACAGGAGGAGCGGATGAGCGACACACAAACACACAGACACAGCCCA 421

QY 422 GTCCAGGAGCCAGTAATGAGAGCCCCAAAAGAAAGAACACAGCAGCTGAAAGTCGGGA 481

QY 482 TCCTACACTGGGACGACAGAGAGATCAGGATACAGCTCAGATCCAGTG----- 536

QY 54 leLeuHisLeuGlySerArgGlnLysLysIleArgIleGlnLeuArgSerGlnValLeuG 74

QY 537 -----CGCACATGAAGTGTATCTGCAAGAGCTGCATCAGTCAAAACACCGGG 585

QY 74 lyArgGluMetArgAspMetGluGlyAspLeuGlnGluLeuHisGlnSerAsnThrGlyA 94

QY 586 ATAAATCTGGATTTGGTTCGCGGCTCAAGGTGAAGATAATACC 629

QY 94 sPlysSerGlyPheGlyPheArgArgGlnGlyGluAspAsnThr 108

RESULT 4

AAB76870

ID AAB76870 standard; Protein: 108 AA.

XX AAB76870;

AC

DT 12-APR-2001 (first entry)

DE Human lung tumour protein related protein sequence SEQ ID NO:793.

XX

KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;

KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;

KW cytostatic; antisense inhibition.

XX

OS Homo sapiens.

XX

PN WO200100828-A2.

XX

PD 04-JAN-2001.

XX

PF 30-JUN-2000; 2000WO-US18061.

} n ✓





30-MAR-2001; 2001WO-US08631.  
 31-MAR-2000; 2000US-0540217.  
 23-AUG-2000; 2000US-0649167.  
 (HYSE-) HYSEQ INC.  
 Drmanac RT, Liu C, Tang YT;  
 WPI; 2001-639362/73.  
 N-PSDB; AAS69484.  
 New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -  
 Claim 20: SEQ ID No 35656; 103pp; English.  
 The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.  
 Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 Sequence 112 AA;  
 Alignment Scores:  
 Pred. No.: 4.73e-16 Length: 112  
 Score: 244.00 Matches: 50  
 Percent Similarity: 70.5% Conservative: 10  
 Best Local Similarity: 58.8% Mismatches: 24  
 Query Match: 17.68% Indels: 1  
 DB: 22 Gaps: 0  
 US-09-658-824-808 (1-781) x ABG05297 (1-112)  
 QY 429 GACCCAGTAAGTGGAGAGCCCAAGAGAGACAGCAGTGAAGTCGGATCCTACA 488  
 Db 29 GLUPROGLYASPLGluProGlnGlnGluGluProThrGluSerArgAspProAla 48  
 QY 489 CCTGGCGACGACAGCAAGATCAGGTGAGATCCAGTCCAGTCGGCGACATGGAA 548  
 Db 49 PROGLYGLNGluArgLysGluaspGlnGlySerAlaLysThrGlnValProaspLyuGlu 68  
 QY 549 GGTGATCTCGAGAGCTGCATCAGTCAACACCCGGGATTAATCTGGATTTGGGTCCGG 608  
 Db 69 AlaAspLyuGlnGluSerGlnSerLysThrGlyGlyGlyCysGlyAsnGlyProasp 88  
 QY 609 CGTCAAGGTGAAGATAATACCTAAAGACAGACACTGTAATGCCAGACAGCGTGAACA 668  
 Db 89 AspGlnGly-LysIleLeuProLysSerGluGlnPhelysMetProGluGlyGlyAsp 108  
 QY 669 GCAACACCAAGTT 681  
 Db 108 gGlnProGlnVal 112



XX WPI: 2001-639362/73.  
DR N-PSDB; AAS91235.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX  
PS Claim 20; SEQ ID NO 57407; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 249 AA;

Alignment Scores:  
Pred. No.: 5.84e-14 Length: 249  
Score: 224.50 Matches: 55  
Percent Similarity: 60.19% Conservative: 10  
Best Local Similarity: 50.93% Mismatches: 33  
Query Match: 16.27% Indels: 10  
DB: 22 Gaps: 1

US-09-658-824-808 (1-781) x ABG27048 (1-249)

QY 399 ACAACACAGAACCCACAGCCAGTCCCGAGGAGCCCGATGAGAGCCCAAAAGAA 458  
Db 61 ThrSerPheLeuProThrMetThrGlnGlnGluProSerValProGluProGlnGlu 80  
QY 459 GAACACAGCAGCTGAAGTCCGGATCTACACCTGGCGAGCAGACAGAGATCAGGAT 518  
Db 81 GluProProThrGluSerGlnAspHisThrProGlyGlnLysArgGluAspGlnGly 100  
QY 519 ACAGCTGAGATCCAGTCCGCGACATGGAGGTGATCTGCAAGAGCTGCATCAGTCAAC 578  
Db 101 AlaAlaGluIleGlnValProAsnLeuGluAlaAspLeuGlnGluSerGlnSerLys 120  
QY 579 ACCGGGATAAATCTGGATTGGTTCGCGCGTCCAGAGTGAAGATAATACCTAAACAGGA 638  
Db 121 ThrGlyAspGluCysGlyAspSerProAspValGlnGly-LysIleLeuProLysSerG 140  
QY 639 ACATGTAAATGCCAGAA-----GCAGGTGAAGAGCA 671  
Db 140 uGlnPheLysMetProGluGlyGluHisLeuArgAspValIleHisAlaCysGluGlu 160  
QY 672 ACCAAGTAAATGAAGACA 693  
Db 160 uGluProValSerMetThrThr 167

RESULT 10  
AAM41374  
ID AAM41374 standard; Protein; 106 AA.  
XX

AAM41374;  
22-OCT-2001 (first entry)  
Human polypeptide SEQ ID NO 6305.  
Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
peripheral nervous system; neuropathy; central nervous system; CNS;  
Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;  
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
leukaemia.  
XX Homo sapiens.  
OS WO200153312-A1.  
PN 26-JUL-2001.  
XX 26-DEC-2000; 2000WO-US34263.  
XX 21-JAN-2000; 2000US-0488725.  
XX 25-APR-2000; 2000US-052317.  
XX 09-JUL-2000; 2000US-0598042.  
XX 19-JUL-2000; 2000US-0620312.  
XX 03-AUG-2000; 2000US-0653450.  
XX 14-SEP-2000; 2000US-0662191.  
XX 19-OCT-2000; 2000US-0693036.  
XX 29-NOV-2000; 2000US-0727344.  
(HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Wang J, Jang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
DR WPI: 2001-442253/47.  
DR N-PSDR; AAI60530.  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
XX such as central nervous system injuries -  
PS Example 2; SEQ ID NO 6305; 10078pp; English.  
XX The invention relates to human nucleic acids (AA157798-AA161369) and  
XX the encoded polypeptides (AAM38642-AA42213) with nootropic,  
XX immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX in gene therapy. A composition containing a polypeptide or polynucleotide  
XX of the invention may be used to treat diseases of the peripheral nervous  
XX system, such as peripheral nervous injuries, peripheral neuropathy and  
XX localised neuropathies and central nervous system diseases, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX utilisation of the activities such as: immune system suppression,  
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
XX assays for receptor activity, arthritis and inflammation, leukaemias and  
XX C.N.S disorders.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification.  
SQ Sequence 106 AA;  
Alignment Scores:  
Pred. No.: 1.23e-13 Length: 106  
Score: 220.00 Matches: 47  
Percent Similarity: 68.83% Conservative: 6  
Best Local Similarity: 61.04% Mismatches: 23  
Query Match: 15.94% Indels: 1  
DB: 22 Gaps: 0  
US-09-658-824-808 (1-781) x AAM41374 (1-106)



```

PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-Nov-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK52902.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 328; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM8020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 117 AA;

Alignment Scores:
Pred. No.:      2.1e-06      Length:      117
Score:          148.50       Matches:    36
Percent Similarity: 57.78%   Conservative: 16
Best Local Similarity: 40.00% Mismatches:    32
Query Match:      10.76%   Indels:     6
DB:               22        Gaps:      1

US-09-658-824-808 (1-781) x AAM79769 (1-117)
OY 429 GAGCCCCAGTAAATGGAGACGCCCAAAAAGAACGACGAGCTGAAGTCGGGATCCTACA 488
Db 29 GluProThrAspGluGluProLysGlnGlyAlaMetValArgProLysSerArgAsnProThr 48
OY 489 CCTGGGCAGACAGACAGAAGATCATCAGGATACAGCTGAGATCCCAGTGCGCCACATGGAA 548
Db 49 ProAspGlnLysArgGluAspSergGlySerAlaAla***AspPheLysTrpProGlu 68
OY 549 GGTGATCTGCAAGAGCTCATCAGTCAACACCACCCGGGATAATCTGGA----- 596
Db 69 ProGlyLysProIlePheGlnGlyAlaMetValArgProLysThrGlyGlyCysGlyCys 88
OY 597 ---TTTGGTTC CGCGCTCAAGGTGAAGATAATACTAAAGAGGAACTGTAAAAATGCC 653
Db 89 GluGlyGlyTyr***CysGlnGlyGluAspSer-ProLysAlaGluHisPheLeuMetPr 108
OY 654 AGAAGCAGGTGAACAGCAACCAACCAAGTT 681
Db 108 oGluAlaGlyGluGlyLysSerGlnVal 117

RESULT 13
AAY83158
ID AAY83158 standard; Protein; 102 AA.
XX
AC AAY83158;
XX
```

QY 561 GAGCTGCATCAGTCAAAACACCGGGGATAAAATCTGGATTTCGGGCTCAAGGTGAA 620  
|||::: ||| ::| ||| ||||| ||| |||  
Db 65 GluMetAspLeuGluLysThrArgSerGluArgGlyAsp-GlySerAspValLysGluLys 84  
| ||| ::| ||| ||||| ||| |||  
QY 621 GATAATACCTAAAGAGGAACACTGTAAATGCCAGAGCAGGTGAAGAGCAACCA 675  
| ||| ::| ||| ||||| ||| |||  
Db 84 sThrProProAsnProLysHisAlaLysThrLysGluAlaGlyAspGlyGlnPro 102  
RESULT 14  
AA152998  
ID AAY52998 standard; Protein; 102 AA.  
XX  
AC AAY52998;  
XX  
DT 29-FEB-2000 (first entry)  
XX  
DE Human secreted protein clone bn365\_53 protein sequence SEQ ID NO:2.  
XX  
KW Human; secreted protein; nutritional; cytokine; cell proliferation;  
KW differentiation; immune stimulating; vaccine; suppression;  
KW haematopoiesis regulation; tissue growth; activin; inhibin;  
KW chemotactic; chemokinetic; haemostatic; thrombolytic; receptor;  
KW ligand; anti-inflammatory; cadherin; tumour invasion suppressor;  
KW tumour inhibition; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO9957132-A1.  
XX  
PD 11-NOV-1999.  
XX  
PF 07-MAY-1999; 99WO-US09970.  
XX  
PR 07-MAY-1998; 98US-0084564.  
PR 02-JUN-1998; 98US-0087645.  
PR 22-JUL-1998; 98US-0093712.  
PR 31-JUL-1998; 98US-0094935.  
PR 10-AUG-1998; 98US-0095880.  
PR 11-AUG-1998; 98US-0096068.  
PR 06-MAY-1999; 99US-0096068.  
XX  
PA (GENY ) GENETICS INST INC.  
XX  
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Bowman MR;  
PI DiBlasio-Smith E, Widom A;  
XX  
DR WPI; 2000-052937/04.  
DR N-PSDB; AAZ33316.  
XX  
PT New polynucleotides encoding secreted human proteins, derived from  
PT adult placenta, adult retina, fetal brain, fetal  
XX  
PS Claim 9; Page 360-361; 492pp; English.  
XX  
CC The present invention describes new human secreted proteins which were  
CC isolated from adult placenta, adult retina, foetal brain, foetal kidney,  
CC adult blood, adult brain, adult thyroid, adult bladder, adult neural  
CC tissue, adult testes, and adult lymph node cDNA libraries. The human  
CC secreted proteins, and the polynucleotides encoding them, are predicted  
CC to have biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals. Suggested activities include nutritional activity, cytokine  
CC and cell proliferation/differentiation activity, immune stimulating  
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, and tumour inhibition  
CC activity. The polynucleotides are also stated to be useful for gene  
CC therapy. AAZ33316 to AAZ33373 encode human secreted proteins, and  
CC AAY52998 to AAY53060 represent human secreted proteins, given in the  
CC present invention.

XX  
SQ Sequence 102 AA;  
Alignment Scores: 2.04e-05 Length: 102  
Pred. No.: 138.50 Matches: 34  
Score: 55.70% Conservative: 10  
Best Local Similarity: 43.04% Mismatches: 31  
Query Match: 10.04% Indels: 4  
DB: 21 Gaps: 1  
US-09-658-824-808 (1-781) x AAY52998 (1-102)  
QY 441 GGAGAGCCCCAAAAGAACACGAGCTGAAAGTCGGGATCCTACACTGGCCAGCAG 500  
||||| |||:::||||||| ::| ||| ||||| ||| |||  
Db 28 GlyGluSerGlnGlnGluProThrArgAspAsnGlnAspIleGluProGlyGlnGlu 47  
QY 501 ACAGAGAAGATCAGGATCAGCTGAGATGCCAGTCCGACATGAAGGTGATCTGCAA 560  
||| ||| ||| ||| ::| ||| ||||| ||| |||  
Db 48 ArgGlu-----GlyThrProIleGluLysValGluGlyAspCysGln 64  
QY 561 GAGCTGCATCAGTCAAAACACCGGGGATAAATCTGGATTTCGGGCTCAAGGTGAA 620  
|||::: ||| ||| ::| ||| ||||| ||| |||  
Db 65 GluMetAspLeuGluLysThrArgSerGluArgGlyAsp-GlySerAspValLysGluLys 84  
QY 621 GATAATACCTAAAGAGGAACACTGTAAATGCCAGAGCAGGTGAAGAGCAACCA 675  
| ||| ::| ||| ||||| ||| |||  
Db 84 sThrProProAsnProLysHisAlaLysThrLysGluAlaGlyAspGlyGlnPro 102  
RESULT 15  
AA12447  
ID AAY12447 standard; Protein; 89 AA.  
XX  
AC AAY12447;  
XX  
DT 17-JUN-1999 (first entry)  
XX  
DE Human 5' EST secreted protein SEQ ID NO:478.  
XX  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition.  
XX  
OS Homo sapiens.  
XX  
PN WO9906548-A2.  
XX  
PD 11-FEB-1999.  
XX  
PF 31-JUL-1998; 98WO-IB01222.  
XX  
PR 01-AUG-1997; 97US-0905135.  
XX  
PS (GEST ) GENSET.  
XX  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
XX  
DR WPI; 1999-153778/13.  
DR N-PSDB; AAX41280.  
XX  
PT New nucleic acids encoding human secreted proteins - obtained from  
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
PT kidney, lung, umbilical cord, placenta and colon tissue  
XX  
PS Claim 27; Page 778; 824pp; English.  
XX  
CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for  
CC human secreted proteins, and encode the proteins given in AAY12261 to  
CC AAY12514, respectively. The proteins given represent the signal peptide  
CC and an N-terminal fragment of a secreted protein. The nucleic acid



KW cytotoxic T lymphocyte; immune response; antibody; drug delivery;  
XX immunoconjugate.  
OS Homo sapiens.  
XX  
PN WO200012706-A1.  
XX  
PD 09-MAR-2000.  
XX  
PF 31-AUG-1999; 99WO-US20046.  
XX  
PR 01-SEP-1998; 98US-0098993.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA  
PI Pastan I, Brinkmann U, Vasmatazis G, Lee B;  
XX WPI; 2000-237869/20.  
DR  
XX  
XX Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T  
PT lymphocyte response and for raising antibodies which can be used to  
PT detect the presence of PAGE-4 in cell samples or body tissues  
XX  
PS Disclosure: Figure 1b; 63pp; English.  
XX  
CC PAGE-4 is a gene preferentially expressed in normal male and female  
CC reproductive tissues e.g. prostate, testis, fallopian tube, uterus  
CC and placenta, as well as in prostate cancer, testicular cancer and  
CC uterine cancer. This expression pattern makes it a target for  
CC diagnosis and for vaccine based therapy of such neoplasms.  
CC An isolated PAGE-4 peptide which induces a cytotoxic T  
CC lymphocyte response when bound to a major histocompatibility complex  
CC (MHC) class I molecule or the isolated PAGE-4 protein can be used in  
CC immunogenic compositions to raise a cytotoxic T lymphocyte response  
CC against cells expressing PAGE-4 including cancer cells of the  
CC prostate, uterus and testis. The nucleic acids encoding PAGE-4 or  
CC PAGE-4 peptide fragments can also be used in these compositions.  
CC Antibodies against PAGE-4 and its peptide fragments can be used in  
CC detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell  
CC samples or body tissues. The presence of PAGE-4 in tissues which are  
CC not related to reproduction can be indicative of the spread of  
CC cancerous reproductive tissue. PAGE-4 can also be used to raise  
CC antibodies which are then used as the targeting group of  
CC immunoconjugates comprising toxins used in therapeutic applications.  
CC This has applications for drug delivery systems. This sequence is  
CC of the PAGE3 polypeptide which shares sequence similarity with  
CC PAGE-4.  
XX  
SQ Sequence 79 AA:  
  
Alignment Scores:  
Pred. No.: 0.00137 Length: 79  
Score: 120.00 Matches: 32  
Percent Similarity: 53.25% Conservative: 9  
Best Local Similarity: 41.56% Mismatches: 33  
Query Match: 8.70% Indels: 3  
DB: 21 Gaps: 1  
  
US-09-658-824-808 (1-781) x AAY83169 (1-79)  
QY 453 AAGAAGAACGACGCTGAAAGTCGGGATCCTACACCTGGCGCAGCAGACAGAAAGAT 512  
Db 6 LysThrAlaProIleGluSerGlnAspTyrThrProGlyGln-----GluArgASP 23  
QY 513 CAGGATACAGTGCAGTCCAGTCGCGCAGCATGGAGGTGATCTGCAAGAGTGCATCAG 572  
Db 24 GluGlyAlaLeuAspPheGlnValProSerLeuAlaAlaTyrLeuTrpGluLeuThrArg 43  
QY 573 TCAACACCGGGGATAATCTGGATTTCGGGTCCGGCGTCAAGGTGAAGATAATACCTAA 632  
Db 44 ProLysThrGlyGlyGluArgGlyAspGlyProAsnValysGlyGluSer-LeuProAs 63  
QY 633 AGAGGAACACTGTAAATGCCAGACGAGTGAAGACCAACCAAGTT 681

Db 63 nLeuGluProValLysIleProGluAlaGlyGluGlyGlnProSerVal 79  
RESULT 18  
AAB07749  
ID AAB07749 standard; Protein; 146 AA.  
XX  
AC AAB07749;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE A human cancer-associated protein-3 (CAP-3).  
XX  
KW Human; cancer-associated protein; CAP; CAP-1; CAP-2; CAP-3;  
KW cell proliferation; arteriosclerosis; cirrhosis; psoriasis; hepatitis;  
KW cancer; leukaemia; melanoma; autoimmune disorder; inflammatory disorder;  
KW acquired immunodeficiency syndrome; AIDS; anaemia; asthma;  
KW Crohn's disease; multiple sclerosis; microbial infection.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 19  
FT /note= "potential casein kinase II phosphorylation site"  
FT Modified-site 23  
FT /note= "potential casein kinase II phosphorylation site"  
FT Modified-site 41  
FT /note= "potential casein kinase II phosphorylation site"  
FT Modified-site 71  
FT /note= "potential casein kinase II phosphorylation site"  
FT Modified-site 81  
FT /note= "potential casein kinase II phosphorylation site"  
FT Modified-site 113  
FT /note= "potential casein kinase II phosphorylation site"  
FT Modified-site 136  
FT /note= "potential casein kinase II phosphorylation site"  
XX  
WO200043508-A2.  
XX  
PN 27-JUL-2000.  
XX  
PD 21-JAN-2000; 2000WO-US01565.  
XX  
PR 22-JAN-1999; 99US-0183027.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Hillman JL, Yue H, Tang YT, Azimzai Y;  
XX  
DR WPI; 2000-482911/42.  
DR N-PSDB; AAA59322.  
XX  
XX Isolated nucleic acids encoding human cancer-associated proteins,  
PT useful for treating cancers and autoimmune/inflammatory disorders -  
XX  
PS Claim 1; Fig 3A-B; 89pp; English.  
XX  
CC The present sequence represents a human cancer-associated protein (CAP).  
CC The specification describes CAP-1, CAP-2 and CAP-3. CAP polynucleotides  
CC and polypeptides may be used in the prevention, treatment and diagnosis  
CC of diseases associated with inappropriate CAP expression. For example,  
CC they may be used to treat disorders associated with decreased CAP  
CC expression such as disorders of cell proliferation  
CC (e.g. arteriosclerosis, cirrhosis, psoriasis and hepatitis), cancers  
CC (e.g. leukaemia, melanoma and cancers of the breast lung and prostate),  
CC autoimmune/inflammatory disorders (e.g. acquired immunodeficiency  
CC syndrome (AIDS), anaemia, asthma, Crohn's disease and multiple  
CC sclerosis), and microbial infections. The CAP polypeptides may be used  
CC as antigens in the production of antibodies against CAP and in assays  
CC to identify modulators (agonists and antagonists) of CAP expression and  
CC activity.  
XX  
SQ Sequence 146 AA;





Db 341 euSerProGlnProAlaSerProAlaGluGluAspLysMetProProTyrAspGluG 361  
Qy 499 AGACAGAGAGATCAGGATACA-----GCTGAGA 528  
Db 361 lInThrGlnAlaPheIleAspAlaAlaGlnGluAlaArgAsnLysPheGluGluAlaGluA 381  
Qy 529 TCCAGTCCGCGACATGAAGGTGATCTGCAAGAGCTGCATCAGTCAACACCGGGGATA 588  
Db 391 rgSerLeuLysAspMetGluGluSerIleArgAsnLeuGluGln-----G 396  
Qy 589 AATCGGATTTGGTCCGCGCTCAAGGTGAA 620  
Db 396 luIleSerPheAspPheGlyProAsnGlyGlu 406  
RESULT 20  
AAU28042  
ID AAU28042 standard; Protein; 527 AA.  
XX  
AC AAU28042;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Novel human secretory protein, Seq ID No 211.  
XX  
KW Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;  
KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;  
KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;  
KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;  
KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;  
KW gut protection; lung; liver fibrosis; immune deficiency; infection;  
KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;  
KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;  
KW fertility; analgesic; pain; antigen.  
XX  
OS Homo sapiens.  
XX  
PN WO200166589-A2.  
XX  
PD 13-SEP-2001.  
XX  
PF 05-MAR-2001; 2001WO-US04942.  
XX  
PR 07-MAR-2000; 2000US-0519705.  
PR 19-MAY-2000; 2000US-0574454.  
PR 17-JUN-2000; 2000US-0596193.  
PR 14-JUL-2000; 2000US-0616847.  
PR 19-SEP-2000; 2000US-0665363.  
PR 20-OCT-2000; 2000US-0693267.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;  
PI Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;  
XX  
XX WPI: 2001-589934/66.  
XX N-PSDB: AAS44942.  
XX  
PT Novel polypeptides and nucleic acids obtained from cDNA libraries  
PT prepared from various human tissues, for diagnosis and treatment of  
PT cancer, neurological, inflammatory, and autoimmune disorders -  
XX  
XX Example 3; SEQ ID No 211; 107pp; English.  
XX  
XX The invention relates to novel isolated human secreted polypeptides (I)  
XX and polynucleotides (II). (I) and (II) are useful for treating  
XX inflammatory conditions such as arthritis, nephritis, Crohn's disease,  
XX ischaemia-reperfusion injury, shock, sepsis, immune responses, and is  
XX involved in increasing haematopoiesis, stem cell survival, bone growth  
XX and remodeling. (I), (II) and modulators of (II) are useful for  
XX prophylaxis or treatment of one or more cancers. (II) is also useful for  
XX creating transgenic animals useful for studying the in vivo activities of  
XX the polypeptide as well as for studying modulators of the polypeptides.  
XX (I) induces the proliferation of neural cells and regeneration of nerve

CC and brain tissue and is useful for the treatment of central and  
CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,  
CC Parkinson's disease, Huntington's disease, and amyotrophic lateral  
CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic  
CC activity, regulation of haematopoiesis and is useful for treating myeloid  
CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia  
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
CC tissue growth, and in tissue repair, healing of burns, incisions,  
CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative  
CC disorders, for periodontal disease. Furthermore, (I) is also useful for  
CC gut protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues, various immune deficiencies and  
CC disorders including severe combined immunodeficiency (SCID), bacterial or  
CC fungal infections, autoimmune disorders e.g. multiple sclerosis,  
CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic  
CC reactions and conditions, such as asthma or other respiratory problems.  
CC In addition, (I) affects biorhythms or circadian cycles of rhythms,  
CC fertility, metabolism, catabolism, anabolism, storage or elimination of  
CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides  
CC analgesic effects or other pain reducing effects, immunoglobulin like  
CC activity and can act as an antigen in a vaccine composition to raise an  
CC immune response. AAU28020-AAU28395 represent novel human secreted protein  
XX amino acid sequences of the invention.  
XX  
SQ Sequence 527 AA;  
Alignment Scores:  
Pred. No.: 0.0697 Length: 527  
Score: 106.00 Matches: 52  
Percent Similarity: 37.74% Conservative: 28  
Best Local Similarity: 24.53% Mismatches: 76  
Query Match: 7.68% Indels: 56  
DB: 22 Gaps: 8  
US-09-658-824-808 (1-781) x AAU28042 (1-527)  
Qy 52 TCCGCTACTGAGACACGGGGTAGTCCACAGCAGATCCAACTGGGAGTTGAAGTGTG 111  
Db 230 SerValThrGluLeu--GlnThrHisProGluLeuAspThrAspGlyAspGlyAlaLeu 248  
Qy 112 AGTGACAGAGTGAA-----GAGGAACACGACGGCTTC 141  
Db 249 SerGluAlaGluAlaGlnAlaLeuLeuSerGlyAspThrGlnThrAspAlaThrSerPhe 268  
Qy 142 CGGAGGGTTGTGGTCAGTGACTCAGAGTGAGAGGCCCTCGAAGTCGTGCTCCCTCTC 201  
Db 269 TyrAspArgValTTPAlaAla-IleArgAspLysTyrArgSerGlu-----283  
Qy 202 ATCGGTGCCACGCCCATGGACCTTCTGTCTCAGCGCCCACTAAGTAGGAGG---AA 258  
Db 284 -----AlaLeuProThrAspLeuProAlaProSerAlaProAspLeuThrGluProLy 301  
Qy 259 GGAGGGCCGAGAGTAGTGGGGGCTCAGGCGAAGCTGGGTGCTGTTGGGGTATCCGAGT 318  
Db 301 sGluGluGln-----304  
Qy 319 CCCAGAGACACTGGACCCCGACAGAGATTTGAGACTCCCCACAGCGGACAGGACAG 378  
Db 305 -ProProValProSerSerProThrGluGlu-----GluGluGluGluGluGlu 321  
Qy 379 GGAGCGCATGACGACACACACAGAACACAGACAGTCCACAGGAGCCAGTA 438  
Db 321 uGluGluAlaGluGluGluGluGluGluGluGluAspSerGluGluAlaProProProL 341  
Qy 439 ATGGAGAGCCCCAAAAGAACACACAGCAGCTGAAGTCGGGATPCTTACACCTGGGAGC 498  
Db 341 euSerProGlnProAlaSerProAlaGluGluAspLysMetProProTyrAspGluG 361  
Qy 499 AGACAGAGAGATCAGGATACA-----GCTGAGA 528  
Db 361 lInThrGlnAlaPheIleAspAlaAlaGlnGluAlaArgAsnLysPheGluGluAlaGluA 381  
Qy 529 TCCAGTCCGCGACATGAAGGTGATCTGCAAGAGCTGCATCAGTCAACACCGGGGATA 588

DB  
OY

132 AGAGGAACCAAGCAGCTTCCGGAGGGTTGTGTCAGTGACATCATCGA...  
123 AGAGGAACCAAGCAGCTTCCGGAGGGTTGTGTCAGTGACATCATCGA... SerAsnGlyGluAspAlaAla 162

150 ArgGlyValPseSer...  
150 ArgGlyValPseSer...  
db db

[illegible]

Db

QY 252 ----GlyArgGlyGlyArgGlySerGlyArg.  
251 ||||  
179 -----GlyArgGlyPheGlyArgGlySerGlyArgGATT 355  
-----CGTCCACCCCGACAGAGATT

5' GCTGGGGTGGTGTGGGGGTATCCGAGTCCAGAGGCT...  
3' ...

195 Ser-----TyrGlyTyrASPAlaAsmo.100  
:::  
195 Ser-----TyrGlyTyrASPAlaAsmo.100

351 CTGGACTCCCGACGGGACCCAGGAGG...  
 351 CTGGACTCCCGACGGGACCCAGGAGG...  
 351 CTGGACTCCCGACGGGACCCAGGAGG...

Db 212 GluAspThr-----ASNinPhe-----  
-----CCGCGCTAATGGAGAGCCCCAAAGAAGAACCCAGCAGCT

[illegible]

471 GAAAGTCGGG...  
QY |||||---valThrSerGluAsnLysASPAsnLys...

Db

[illegible]

D0  
265  
501

TCTGGATTGGGTTCCGGCGTC AAGGTGAAGATATACCTT...  
... |||::||| ||| |||:::  
... |||::||| ||| |||:::

PheGluLysM

282 rLeuGluAlaLeuLysAsnAlaGluArgLysValr  
--AAGTTTAATGAAGACAGCTC

651 GCCAGAGCAGGTGAAGACCAAC

302 etGlnLeuValAspLysLysAsnAsp-1

699 AAACAACGCAAG 710  
QY

Db 322 ArgGlnArgLys 323

RESULT 22  
 1AW7Q958  
 ... protein: 578 AA.

ID  
AAW79958 standard  
YX

AC AAW/99307  
XX 1000 (first entry)

[illegible]

XX progression elevated gene-3; PEG-3; human, melanoma  
XX angiogenesis; metastasis; lung cancer;

DNA replar; cancer;  
KW cervix cancer; prostate  
v12 brain cancer; glioblastoma;  
v13 neuroblastoma; diagnosis;

KW  
XX  
XX

OS XX  
XX  
HOMING SOFT  
-00015-A]

PN  
XX  
01-007-1998.

XX 98WO-US05793.  
20-MAR-1998;

XX 21-MAR-1997; 97US-0821818.

XX (GENQ-) GENQUEST INC.  
PA

PA (UYCO ) UNIV COLUMBIA NEW YORK.

PI Fisher PB, Goldstein NI, Su Z, Zhang N;

XX WPI; 1998-557025/47.

DR N-PSDB; AAV65765.

XX New isolated Progression Elevated Gene-3 - used to develop products

PT for e.g. modulating DNA damage and repair pathways, cancer

PT progression or oncogene mediated transformation and angiogenesis.

XX Claim 17; Fig 13A-C; 225pp; English.

PS This polypeptide comprises human progression elevated gene-3 (PEG-3)

XX protein. The amino acid sequence was deduced from an isolated cDNA

CC clone (see AAV65765). PEG-3 expression (1.5 and 2.8 kb transcripts)

CC is elevated in most human tumour cell lines. The amount of PEG-3

CC in a cell is also an indicator of DNA damage in that cell. Cells

CC in which a reporter gene is under control of the promoter or

CC regulatory region (see AAV65766) of the rat PEG-3 gene can be used in

CC methods for identifying agents that modulate PEG-3 expression or the

CC ability of PEG-3 to induce progression, or for determining whether

CC an agent is capable of inhibiting DNA damage and repair pathways,

CC cancer progression or oncogene-mediated transformation. Compounds

CC that induce DNA damage or which regulate angiogenesis can also be

CC identified using such cells. Transgenic animals and vaccines

CC comprising PEG-3 polypeptides and an immune response enhancer are

CC also claimed. Antibodies (especially monoclonal) to the PEG-3 may

CC be used to determine tumour progression of a cell. Cells can be

CC protected from chemotherapeutic damage by inhibiting or eliminating

CC the expression of PEG-3 in the cells. The methods can be applied

CC to a progression phenotype comprising anchorage-independent growth,

CC tumorigenesis, angiogenesis or metastasis, to melanoma, brain,

CC cervical, prostate, lung or colorectal cancer, neuroblastoma or

CC glioblastoma (all claimed).

XX Sequence 578 AA;

SQ

Alignment Scores:

Pred. No.: 0.145 Length: 578

Score: 103.00 Matches: 49

Percent Similarity: 37.99% Conservative: 19

Best Local Similarity: 27.37% Mismatches: 67

Query Match: 7.46% Indels: 44

DB: 19 Gaps: 9

US-09-658-824-808 (1-781) x AAV79958 (1-578)

QY 112 AGTGTAGAGTGAAGAGAACACGAGCGCTTCCGGAGGGTTGTGTGCTCAGTGCAGAGT 171

DB 220 ThrAspAsnLysAlaGluProSerGlySerHisSerArgPheTrpGluTyrHisThrArg 239

QY 172 GAGAAGCCCTCGAAGTCGTGCTCCCTCATGCGGTGCCAGCCCATGGACCTTCTTGT 231

DB 240 GluArgProLysGlnGluGlyGluThrLysProGluGlnHisArgAlaGlyGlnSerHis 259

QY 232 CTCGTCACGCCATACTAGGAGGAGGAGGCCCGCAGG-----AGTGGAGGG 279

DB 260 ProCysGlnAsnAlaGluSerGluGluGlyProGluThrSerValCysSerGlySer 279

QY 280 GCTCAGCGAAGCTGGGTGCTGTTGGGGTATCCGAGTCCAGAACACCTGGAAACCCC 339

DB 280 AlaPheLeuLys-----AlaTrpValTyr-Arg-----ProGlyGluAs 292

QY 340 GACAGAGATTCGTGACTCCCA-----GACGGGACAGGAGGAGCGGATGAGCGA 393

DB 292 pThrGluGluGlu-GluAspProAspLeuAspSerAlaGluAspThrAla----- 309

QY 394 CACACAAACACACAGAACACACAGC----- 419

DB 310 --HisThrCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProG 329

QY 420 -----CAGTCCCAGGAGCCCAAGTAATGGAGAG---CCCAAAAGAGAACACGAGCAGT 471

Db 329 lyGluAspThrGluGluGluAspGlyAspTrpAspSerAlaGluGluAspAlaG 349

QY 472 AAAGf---CGGATCCT-----ACACCTGGGC 495

Db 349 InSerCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyG 369

QY 496 AGCAGACAGAGAGATCAGGATACAGCTCAGATCCAGTCCGCGGACATGAA 548

Db 369 luAspThrGluGluGluAspSerGluAsnValAlaProValAspSerGlu 386

RESULT 23

AA41104

ID AAY41104 standard; Protein: 578 AA.

XX

AC AAY41104;

XX

DT 17-JAN-2000 (first entry)

XX

DE Human progression elevated gene-3 (PEG-3) amino acid sequence.

XX

KW Progression-elevated gene-3; PEG-3 gene; regulatory region; cell death;

KW cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; human.

XX

OS Homo sapiens.

XX

PN M09949898-A1.

XX

PD 07-OCT-1999.

XX

PF 31-MAR-1999; 99WO-US07199.

XX

PR 31-MAR-1998; 98US-0052753.

XX

PA (UYCO ) UNIV COLUMBIA NEW YORK.

XX

PI Fisher PB;

XX

DR WPI; 1999-591184/50.

XX

N-PSDB; AAZ23029.

XX

Novel vectors useful for studying the progression of cancer -

PS Disclosure; Fig 13A-C; 251pp; English.

XX

CC The invention relates to an inducible progression-elevated gene-3 (PEG-3

CC gene) regulatory region functionally linked to a gene encoding a product

CC that causes or may be induced to cause the death or inhibition of cancer

CC cell growth. A vector of the invention which contains a gene encoding a

CC thymidine kinase or a product which causes the cell to express a

CC specific antigen can be administered along with gancyclovir or acyclovir,

CC or an antibody or fragment to the antigen, respectively, to treat cancer

CC in a subject. The PEG-3 gene is useful for generating new cloning and

CC expression vectors, transfected cells, and for developing methods for

CC cultured growth of such cells. The PEG-3 polynucleotide is also useful as

CC a source of primers and probes to study the progression of cancer, and to

CC detect the presence of the gene. The present sequence represents the

CC amino acid sequence of the human PEG-3 protein.

XX

SQ Sequence 578 AA;

Alignment Scores:

Pred. No.: 0.145 Length: 578

Score: 103.00 Matches: 49

Percent Similarity: 37.99% Conservative: 19

Best Local Similarity: 27.37% Mismatches: 67

Query Match: 7.46% Indels: 44

DB: 20 Gaps: 9

US-09-658-824-808 (1-781) x AAY41104 (1-578)

QY 112 AGTGTAGAGTGAAGAGAACACGAGCGCTTCCGGAGGGTTGTGTGCTCAGTGCAGAGT 171

DB: 112

Db 220 ThrAspAsnLysAlaGluProSerGlySerHisSerArgPheTrpGluThrHisThrArg 239  
QY 172 GAGAAGCCCTCGAAGTCGTCCTCTCATCGGTGCCACGCCCATGGACCTTCTTGT 231  
Db 240 GluArgProLysGlnGluGlyGluThrLysProGluGlnHisArgAlaGlyGlnSerHis 259  
QY 232 CTCGTACAGCCATAACTAGGAGGAGGAGGCCCGAGG-----AGTGGAGG 279  
Db 260 ProCysGlnAsnAlaGluSerGluGlyGlyProGluThrSerValCysSerGlySer 279  
QY 280 GCTCAGCGAAGCTGGGGTCTGTTGGGGTATCCGAGTCCCGAGACACCTGGAACCCC 339  
Db 280 AlaPheLeuLys-----AlaTrpValTyr-Arg-----ProGlyGluAs 292  
QY 340 GACAGAGATTCTGGACTCCCA-----CACGGGACCAGAGGAGCGGCATGACCGA 393  
Db 292 pThrGluGluGlu-GluAspProAspLeuAspSerAlaGluGluAspThrAla----- 309  
QY 394 CACACACAAACACAGAACCCACAGC-----ACACCTGGGC 419  
Db 310 --HisThrCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProG 329  
QY 420 -----CAGTCCCGAGGCCAGTAATGGAGAG---CCCCAAAAGAACACAGCAGCTG 471  
Db 329 lyGluAspThrGluGluGluAspGlyAspTrpAspSerAlaGluGluAspAlaAlaG 349  
QY 472 AAAGT---CGGGATCCT-----ACACCTGGGC 495  
Db 349 InSerCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyG 369  
QY 496 AGCAGACAGAAGATCAGGATACAGTCCAGTCCGAGTCCGCGGCACATGGAA 548  
Db 369 luAspThrGluGluGluAspAspSerGluAsnValAlaProValAspSerGlu 386  
RESULT 24  
ID AAY06514  
ID AAY06514 standard; Protein; 578 AA.  
XX  
AC AAY06514;  
XX  
DT 08-OCT-1999 (first entry)  
XX  
DE Human progression elevated protein-3 (PEG-3).  
XX  
KW Progression elevated protein-3; PEG-3; human; angiogenesis; cancer;  
KW therapy; vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO9937776-A1.  
XX  
PD 29-JUL-1999.  
XX  
PF 26-JAN-1999; 99WO-US01623.  
XX  
PR 29-JAN-1998; 98US-0072941.  
PR 26-JAN-1998; 98US-0072469.  
XX  
PA (GENQ-) GENQUEST INC.  
XX  
PI Fisher PB, Zhang N;  
XX  
DR WPI; 1999-458694/38.  
DR N-PSDB; AAX87383.  
XX  
PT Modulation of angiogenesis by altering the expression and/or  
PT activity of a progression-associated protein, especially for cancer  
PT treatment  
XX  
PS Claim 1; Page 74-76; 81pp; English.  
XX  
CC The present sequence represents human progression elevated protein-3  
CC (PEG-3), as deduced from a cDNA clone (see AAX87383) isolated from a

CC MCF-7 cDNA library. PEG-3 represents a novel member of the  
CC gad34/MyD116 gene family. Human PEG-3 is generally expressed in  
CC cells that are in progression, including most tumour cell lines. A  
CC claimed method for modulating angiogenesis in an organism comprises  
CC administering an agent that alters expression and/or activity of a  
CC PEG-3 protein, such as an antisense polynucleotide or antibody.  
CC Also claimed is a method for determining whether an agent  
CC modulates angiogenesis, where the candidate agent is present  
CC within a combinatorial small molecule library. Vaccines and  
CC pharmaceutical compositions comprising such compounds are also  
CC provided and may be used to prevent angiogenesis, especially  
CC related to cancer cell progression.  
XX  
SQ Sequence 578 AA;  
Alignment Scores: 0.145 Length: 578  
Pred. No.: 103.00 Matches: 49  
Score: 37.99% Conservative: 19  
Percent Similarity: 27.37% Mismatches: 67  
Best Local Similarity: 7.46% Indels: 44  
Query Match: 20 Gaps: 9  
DB:  
US-09-658-824-808 (1-781) x AAY06514 (1-578)  
QY 112 AGTGAGAGTGAAGAGAACACGAGCGCTTCCGGAGGTTGTGTGTCAGTCACTCAGAGT 171  
Db 220 ThrAspAsnLysAlaGluProSerGlySerHisSerArgPheTrpGluThrHisThrArg 239  
QY 172 GAGAAGCCCTCGAAGTCGTCCTCTCATCGGTGCCACGCCCATGGACCTTCTTGT 231  
Db 240 GluArgProLysGlnGluGlyGluThrLysProGluGlnHisArgAlaGlyGlnSerHis 259  
QY 232 CTCGTACAGCCATAACTAGGAGGAGGAGGCCCGAGG-----ACTGGAGG 279  
Db 260 ProCysGlnAsnAlaGluSerGluGlyGlyProGluThrSerValCysSerGlySer 279  
QY 280 GCTCAGCGAAGCTGGGGTCTGTTGGGGTATCCGAGTCCCGAGACACCTGGAACCCC 339  
Db 280 AlaPheLeuLys-----AlaTrpValTyr-Arg-----ProGlyGluAs 292  
QY 340 GACAGAGATTCTGGACTCCCA-----CACGGGACCAGGAGGAGCGGCATGACCGA 393  
Db 292 pThrGluGluGlu-GluAspProAspLeuAspSerAlaGluGluAspThrAla----- 309  
QY 394 CACACACAAACACAGAACCCACAGC-----ACACCTGGGC 419  
Db 310 --HisThrCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProG 329  
QY 420 -----CAGTCCCGAGGCCAGTAATGGAGAG---CCCCAAAAGAACACAGCAGCTG 471  
Db 329 lyGluAspThrGluGluGluAspGlyAspTrpAspSerAlaGluGluAspAlaAlaG 349  
QY 472 AAAGT---CGGGATCCT-----ACACCTGGGC 495  
Db 349 InSerCysThrThrProHisThrSerAlaPheLeuLysAlaTrpValTyrArgProGlyG 369  
QY 496 AGCAGACAGAAGATCAGGATACAGTCCAGTCCGAGTCCGCGGCACATGGAA 548  
Db 369 luAspThrGluGluGluAspAspSerGluAsnValAlaProValAspSerGlu 386  
RESULT 25  
ID AAB95541  
ID AAB95541 standard; Protein; 1127 AA.  
XX  
AC AAB95541;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:18149.  
XX  
KW Human; ; imer; detection; diagnosis; antisense therapy; gene therapy.  
XX



FT	Region	760..840	
FT		/note= "Gln, Glu-rich region"	
XX	WO200029626-A1.		
XX			
XX	25-MAY-2000.		
PD			
XX			
XX	19-NOV-1999;	99WO-US27508.	
XX			
XX	19-NOV-1998;	98US-0109422.	
XX	21-APR-1999;	99US-0298568.	
XX			
XX	(KIEFF/) KIEFF E D.		
PA	(BALL/) BALLESTAS M E.		
PA	(KAYE/) KAYE K M.		
XX			
XX	Kieff ED, Ballestas ME, Kaye KM;		
XX			
XX	WPI; 2000-387829/33.		
DR	N-PSDB; AAA30290.		
XX			
PT	Treating or preventing a disease associated with rhodino virus		
PT	Infection in a mammal which includes Kaposi's Sarcoma and Primary		
PT	Effusion Lymphoma		
XX			
XX	Disclosure: Fig 7; 70pp: English.		
XX			
CC	The present sequence is the Kaposi's sarcoma-associated herpesvirus,		
CC	(KSHV) latency-associated nuclear antigen (LANA). KSHV is also known		
CC	as Human Herpes Virus 8 (HHV8) and belongs to the rhadino virus, or		
CC	gamma-2 herpes virus class. The LANA protein is necessary for the		
CC	efficient persistence of rhadino virus DNA in mammalian cells. Persistent		
CC	rhadino virus infection is implicated in a variety of diseases e.g.		
CC	Kaposi's Sarcoma (KS), Primary Effusion Lymphoma (PEL) and multicentric		
CC	Castleman's disease. In addition, KS is a common malignancy in HIV		
CC	patients. KSHV persists in host cells in a latent form. One of the few		
CC	genes expressed from the latent viral DNA is LANA. LANA associates with		
CC	both human chromosomes and with the rhadino virus cis-acting element		
CC	(RVCAE), thereby providing a tethering function: the KSHV DNA episome is		
CC	"tied" to the host chromosomes. This allows the viral DNA to persist in		
CC	the host cell. The present sequence may be used to screen and identify		
CC	molecules that inhibit LANA interaction with RVCAE, thereby interfering		
CC	with the latency cycle of this virus. Potential antiviral treatments for		
CC	the above mentioned diseases may therefore be based on LANA deregulation.		
XX			
SQ	Sequence	1162 AA;	
	Alignment Scores:		
	Pred. No.:	0.208	Length: 1162
	Score:	102.50	Matches: 36
	Percent Similarity:	36.0%	Conservative: 21
	Best Local Similarity:	22.7%	Mismatches: 59
	Query Match:	7.43%	Indels: 42
	DB:	21	Gaps: 4
US	-09-658-824-808 (1-781) x AAY96255 (1-1162)		
QY	261 AGGCCGAGGAGTGAGGGCTCAGCGAAGCTGGGTGCTGTTGGGGGTATCCGAGTCC	320	
Db	422 LysLysGluaspGluaspGlyGlyaspGlyAsnLysThrLeuSerIleGlnSer	441	
QY	321 CAGAAGCACTTGAACCC-----CGACAGAAG	347	
Db	442 GlnGlnGlnGlnGluProGlnGlnGlnGlnGluProGlnGlnGlnGlu	461	
QY	348 ATTCTGGACTCCCGACGCGGACGAGGACGGCATGAGCGACACACACAACACA	407	
Db	462 ProLeuGlnGluProGln-----GlnGln	469	
QY	408 GAACCAACAGCCAGTCC-----CAGGAGCCCAAGTAATGGAGAG	446	
Db	470 GluProGlnGlnGlnGluProGlnGlnGlnGluProLeuGlnGluProGlnGlnGlnGlu	489	

WPI; 2000-097142/08.

New methods and compositions for the detection of human herpesvirus -

Claim 2; Page 59-62; 68pp; English.

Sequences AY58480-Y58532 represent immunogenic polypeptides derived from human herpes virus type 8 (HHV8, a gammaherpesvirus). HHV8 plays an important role in the pathogenesis of AIDS-related Kaposi's sarcoma. The invention relates to a novel method of detecting the presence of human herpesvirus 8 in a biological sample using peptides representative of dominant antigenic regions of HHV8. The method comprises contacting one or more isolated, immunogenic HHV8 peptides with an antibody-containing biological sample, and detecting the formation of a complex between the peptide and the antibody. The presence of a peptide-antibody complex indicates the presence of human herpesvirus 8. The detection of HHV8 infection can be used to diagnose AIDS-associated Kaposi's sarcoma. The HHV8-specific antibodies are useful therapeutically when for the passive immunisation of a human against HHV8 infection, thereby reducing HHV8 related disease. The detection assays are highly specific, sensitive and accurate. Early detection and treatment of Kaposi's sarcoma could diminish the severity of symptoms related to AIDS and the sensitive techniques could reduce erroneous characterisations of skin disorders. Previous assays for HHV8 antibodies such as immunofluorescence assays, immunoblots and enzyme immunoassays lack the sensitivity and accuracy needed for reliable diagnosis of Kaposi's sarcoma. Further advantages







KW	transcription activation; DEAD box RNA dependent helicase;
KW	adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.
XX	Homo sapiens.
OS	WO200073467-A1.
PN	07-DEC-2000.
XX	
PD	
XX	25-MAY-2000; 2000WO-US14719.
PF	
XX	27-MAY-1999; 99US-0136620.
PR	
XX	25-MAY-2000; 2000US-0579181.
PR	
XX	(UYSL-) UNIV SAINT LOUIS.
PA	
XX	
PI	Chrivia J, Yaciuk P;
XX	
PI	WPI; 2001-061545/O7.
DR	N-PSDB; AAC89860.
XX	
PT	Snf2 related cAMP regulatory element (CREB) binding protein (CBP)
PT	activator protein, capable of co-activating CREB binding protein,
PT	useful for modulating transcription and for affecting viral infection
XX	
PS	Claim 1; Page 86-94; 103pp; English.
XX	
CC	The present sequence is an Snf2 related CREB (cAMP regulatory element)
CC	binding protein (CBP) activator protein (SRCAP) polypeptide. It has
CC	Arpase activity and is capable of activating transcription. SRCAP
CC	polypeptides are useful for activating transcription in a cell, for
CC	enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated
CC	activation of transcription in a cell, for treating a patient having a
CC	disease involving a function such as insufficient transcription of a
CC	gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent
CC	helicase, adenoviral DBP protein, beta-actin or a nuclear receptor
CC	affected by SRCAP protein. Compounds that modulate SRCAP function, such
CC	as antibodies, antisense molecules, polynucleotides or ribozymes, are
CC	useful for treating diseases mediated by SRCAP-activated transcription
CC	for example, infection by adenovirus, hepatitis C virus, human
CC	immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or
CC	hepatitis B virus.
XX	
SQ	Sequence 2972 AA:
Alignment Scores:	
Pred. No.:	0.291 Length: 2972
Score:	102.50 Matches: 70
Percent Similarity:	42.11% Conservative: 34
Best Local Similarity:	28.34% Mismatches: 81
Query Match:	7.19% Indels: 12
DB:	Gaps: 12
US-09-658-824-808 (1-781) x AAB50363 (1-2972)	
QY	640 GTTCCTTTAGTATTACTTCACCTTGACCGCGGAACCCAATGCAGAT----- 590 
Db	1291 ValProAlaSerAlaLeuAlaSerProPheProSerAlaProAsnProAlaProAlaGln 1310 
QY	589 -----TTATCCCCGGTGTGTTGACTGATGCAGCTCTTGCAATCACCTTCCATGTCG 539 
Db	1311 AlaserLeuAlaPro-----AlaSerSerAlaSerGlnAlaLeuAla 1325 
QY	538 CGCACTGGGATCTCAGCTGATCCCTGATCTTCTGCTGCTGCCAAGTGTAGGATCC 479 :::
Db	1326 ThrProLeuAlaProMetAlaAlaProGlnThrAlaIleLeuAlaProSerProAlaPro 1345 
QY	478 CGACTTTCAGCTGCTGTTCTCTCTTTGGGCTCTCCATTACTGGGCTCTCTGGGACTGG 419 
Db	1346 ProLeuAlaProLeu-----ProvAlLeu-AlaProSerProGl 1358 
QY	418 CTGTGTG---GTTCTGTGTTGTGTGTGCTCATGCGCTCTCTCTGTCCTCGTCT 362         :    :    :    :

CC The present sequence is an Snf2 related CREB (cAMP regulatory element)  
 CC binding protein (CBP) activator protein (SRCAP) polypeptide. It has  
 CC ATPase activity and is capable of activating transcription. SRCAP  
 CC polypeptides are useful for activating transcription in a cell, for  
 CC enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated  
 CC activation of transcription in a cell, for treating a patient having a  
 CC disease involving a function such as insufficient transcription of a  
 CC gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent  
 CC helicase, adenoviral DBP protein, beta-actin or a nuclear receptor  
 CC affected by SRCAP protein. Compounds that modulate SRCAP function, such  
 CC as antibodies, antisense molecules, polynucleotides or ribozymes, are  
 CC useful for treating diseases mediated by SRCAP-activated transcription,  
 CC for example, infection by adenovirus, hepatitis C virus, human  
 CC immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or  
 CC hepatitis B virus.  
 XX  
 CC  
 SQ Sequence 3118 AA;

Alignment Scores:  
 Pred. No.: 0.296 Length: 3118  
 Score: 102.50 Matches: 70  
 Percent Similarity: 42.11% Conservative: 34  
 Best Local Similarity: 28.34% Mismatches: 81  
 Query Match: 7.19% Indels: 62  
 DB: 22 Gaps: 12

US-09-658-824-808 (1-781) x AAB50362 (1-3118)

QY 640 GTTCCCTTTAGTATATCTTCACCTTGACCGCGAACCACCAATCCAGAT----- 590  
 DB 1437 ValProAlaSerAlaLeuAlaSerProPheProSerAlaProAsnProAlaProAlaGln 1456  
 QY 589 -----TTATCCCGGTGTTTCACTGATGCAGCTCTTGCGAGATCACTTCCATGTCG 539  
 DB 1457 AlaSerLeuLeuAlaPro-----AlaSerSerAlaSerGlnAlaLeuAla 1471  
 QY 538 CGCACTGGGATCTCAGCTGTATCTCTCTGCTGCTGCGCCAGGTAGGATCC 479  
 DB 1472 ThrProLeuAlaProMetaAlaAlaProGlnThrAlaLeuAlaProSerProAlaPro 1491  
 QY 478 CGACTTTCAGCTGCTGCTCTCTTTGGGCTCTCCATTAAGTCTGCGGCTCGGACATGG 419  
 DB 1492 ProLeuAlaProLeu-----ProValLeu-AlaProSerProGln 1504  
 QY 418 CTGTGTG---GTTCTGTGTTGTGTGCTGCTCATGCGCTCCCTCTGCTGCTGCTGCT 362  
 DB 1504 YAlaAlaProValLeuAlaSerSerGlnThrProValProValMet----- 1519  
 QY 361 GGGAGTCCAGAAATCTTCTGCTGGGTTCCAGGTGCTTCTGGGATCGGATACCCGCCAAC 302  
 DB 1520 ---AlaProSerSerThrProGlyThrSerLeuAlaSerAlaSerProValPro----- 1536  
 QY 301 AGCACCCAGCTTCGCTGAGCCCTCCACT-----CCTCGGCC 263  
 DB 1537 -AlaProThrProValLeuAlaProSerSerThrGlnThrMetLeuProAlaProValPr 1556  
 QY 262 CTCCTCTCTCCCTAGTATGCGGTGACGAGACAGAGTCCATGGCGTGGACCGCA 203  
 DB 1556 oSerProLeuProSerProAlaSerThrGlnThr-----LeuAlaLeuAlaPro-A 1573  
 QY 202 TGAGAGGACGACACTTCAGGGCTCTCACTCTGAGTCACTG-----ACCACA 152  
 DB 1573 laLeuAlaProThr-LeuGlyGlySerProSerProSerGlnThrLeuSerLeuGlyThrGly 1592  
 QY 151 CAACCCCTCCGAAGCTGCTGTTCTCTCACTCTCACTCTCACTC----- 111  
 DB 1593 AsnProGlnGly--ProPheProThrGlnThrLeuSerLeuThrProAlaSerSerLeuV 1612  
 QY 110 -----ACACTTCACTCCAGTTCGATCGCTGCTGAGCTACCCGCC 68  
 DB 1612 alProThrProAlaGlnThrLeuSerLeuAlaProGlyProProLeuGlyProThrGlnT 1632  
 QY 67 GTGTCTCAGTAGCGAG-----AAAGAATCCAGACCTCAGGAGCCCGAGTCTGC 20

Db 1632 hrLeuSerLeuAlaProAlaProLeuAlaProAlaSerProValGlyPro---AlaP 1651  
 QY 19 CGGCTCACAGC 9  
 Db 1651 roAlaHisThr 1654  
 RESULT 32  
 ABG21565  
 ID ABG21565 standard; Protein: 600 AA.  
 XX  
 AC ABG21565;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #21556.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS85752.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 51924; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences, (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences, ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 600 AA;  
 Alignment Scores:  
 Pred. No.: 0.185 Length: 600  
 Score: 102.00 Matches: 50  
 Percent Similarity: 37.44% Conservative: 29

```
Best Local Similarity: 23.70% Mismatches: 76
Query Match: 7.39% Indels: 56
DB: 22 Gaps: 8

US-09-658-824-808 (1-781) x ABG21565 (1-600)
Qy 52 TCCGCTACTGAGACACGGCGGTAGGTCCACAGCGAGATCCAACTGGGAGTTGAAGTGTG 111
   ||| ||||| : : : ||| : : :
Db 268 SerValThrGluLeu---GlnThrHisProGluLeuAspThrAspGlyAlaLeu 286
   ||| ||||| : : : ||| : : :
Qy 112 ACTGAGAGTGA-----CAGGAACAGCAGCGCTTC 141
   ||| ||||| : : : ||| : : :
Db 287 SerGluAlaGluAlaGlnAlaLeuLeuSerGlyAspThrGlnThrAspAlaThrSerPhe 306
Qy 142 CGGAGGGTGTGGTTCAGTCACTCAGAGTGAGAGGCCCTCGAAGTCGTCGTCCTCTC 201
   ||| ||||| : : : ||| : : :
Db 307 TyrAspArgValTrpAlaAla-1leArgAspLysTyrArgSerGlu----- 321
Qy 202 ATGCGGTGCCACGCCCATGACCTTCTGTCTGTACGCCCATTAAGTGGAGG---AA 258
   ||| ||| : : : ||||| : : :
Db 322 -----AlaLeuProThrAspLeuProAlaProSerAlaProAspLeuThrGluProLy 339
Qy 259 GGAGGGCGGAGGAGTGGAGGGCTCAGCGGAAGCTGGGTGCTGTGGGGGTATCCGAGT 318
   ||| ||| : : : ||||| : : :
Db 339 sGluGluGln----- 342
Qy 319 CCCAGAGCACTGGAAACCCGACAGAGAAGATTCGTGGACTCCCGACGGGACCGAGAG 378
   ||| ||| : : : ||||| : : :
Db 343 -ProValProSerProThrGluGluGluGluGluGluGluGluGluGluGluGluG 362
Qy 379 GGACGGCATGAGCACACACACAAACACAGAACACACAGCCAGTCCCGAGGCCAGTA 438
   ||| ||| : : : ||||| : : :
Db 362 uGluAlaGluGluGluGluGluGlu-AspSerGluValGlnGluGluGlnProLySG 382
Qy 439 AT-----GGAGAGCCCAAGAAAGAACCCAGCAGCTGAAAGTCGGGATC 483
   ||| ||| : : : ||||| : : :
Db 382 luAlaProProLeuSerProProGlnProAlaSerProAlaGluGluAspLysMetP 402
Qy 484 CTACACCTGGCCACGACACACACAGAAAGATCAGGATACA----- 521
   ||| ||| : : : ||||| : : :
Db 402 roProTyAspGluGlnThrGlnAlaPheIleAspAlaAlaGlnGluAlaArgAsnLysP 422
Qy 522 -----GCTGAGATCCCGGCGGACATGGAAGGTGATGTCGAAGAGCTGCATCAGT 573
   ||| ||| : : : ||||| : : :
Db 422 heGluGluAlaGluArgSerLeuLysAspMetGluSerIleArgAsnLeuGluGlnG 442
Qy 574 CAACACCGCGGATAAATCTGGATTGGG 602
   : : : ||| : : :
Db 442 luIleSerPheAsp-----PheGly 448

RESULT 33
AAW87503
ID AAW87503 standard; Protein; 1212 AA.
XX
AC AAW87503;
XX
DT 23-FEB-1999 (first entry)
XX
DE Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA22.
XX
KW Human; N-methyl-D-aspartate receptor; NMDAR2C;
XX
OS NMDA-activated cation-selective ion channel; glutamate receptor.
XX
PN Homo sapiens.
XX
PD US5849895-A.
XX
PF 15-DEC-1998.
XX
PR 20-APR-1994; 94US-0231193.
XX
PR 20-APR-1994; 94US-0231193.
XX
PR 20-APR-1993; 93US-0052449.
XX
```

(SIBI-) SIBIA NEUROSCIENCES INC.  
Daggett LP, Lu C;  
WPI; 1999-069812/06.  
N-PSDB; AAV82909.  
DNA encoding N-methyl-D-aspartate receptor subunit - useful for the assembly of functional glutamate receptor subunits  
Example 3; Columns 253-262; 203pp; English.  
The present sequence represents a human N-methyl-D-aspartate (NMDA) receptor subunit (NMDAR). The nucleic acid sequence does not contain the 366 5'-most nucleotides, by the insertion of 11 nucleotides between the 366 5'-most nucleotides, nor the 15 nucleotides at positions 1300 and 1301, nor the 15 nucleotides, as set forth in AAV82889. The 1960-1974, nor the 1061 3' nucleotides, as set forth in AAV82889. The cDNA sequence is derived from clone NMDA21. The NMDAR subunits contribute to the formation of NMDA-activated cation-selective ion channels. In addition to being useful for the production of NMDA receptor subunit proteins, the nucleic acids are also useful as probes to identify and isolate nucleic acids encoding related receptor subunits. Functional glutamate receptors can be assembled from several NMDA receptor subunit proteins of one type (homomeric) or from combinations of subunit proteins of different types (heteromeric). The present invention also comprises methods for using such receptor subunits to identify and characterize compounds which affect the function of such receptors, e.g. agonists, antagonists and modulators of glutamate receptor function. The invention also comprises methods for determining whether unknown protein(s) are functional as NMDA receptor subunits.

XX SQ Sequence 1212 AA;  
Alignment Scores:  
Pred. No.: 0.337 Length: 1212  
Score: 100.50 Matches: 70  
Percent Similarity: 32.75% Conservative: 23  
Best Local Similarity: 24.65% Mismatches: 95  
Query Match: 7.05% Indels: 96  
DB: 20 Gaps: 12

US-09-658-824-808 (1-781) x AAW87503 (1-1212)  
Qy 677 TGTGGTGTCTTTCACGCTGCTTTCGCAATTTTACAGTGTCTCTTTAGGTATATCTTC 618  
 ||| ||||| : : : ||| : : :  
Db 532 CysGlyCysCysGlyArgTrpSerSerThrThrGlnCys-ProSerArgThrArgGlyAl 551  
Qy 617 ACCTTGACGCCG-----AACCCAAATCCAGA 591  
 ||| ||| : : : ||| : : :  
Db 551 aProProAlaArgSerTrpPheTrpSerGlyProSerLeuLeuSerSerSerProAs 571  
Qy 590 TTTATCCCGGTGTTTGACTGATGCAGC----- 563  
 ||| ||| : : : ||| : : :  
Db 571 pThrArgProThrTrpProSerSerLysSerAsnThrSerThrLeuLeuSerArgAla 591  
Qy 562 -----TCTTGCAGATC 552  
 ||| ||| : : : ||| : : :  
Db 591 rValThrArgSerPheSerGlyLeuLysIleSerThrHisLeuSerAlaSerAlaArgCy 611  
Qy 551 ACCTTCCATGTCGCGCAGCTGATCTCAGCTGTATCTCTGATCTCTCTCTGCTGCTGCC 492  
 ||| ||| : : : ||| : : :  
Db 611 sProThrAlaAlaArgSerGlyThrSerAlaVal-----ThrThrValThrCysThrPr 629  
Qy 491 AGGTGTAGGATCCGCACTTTCAGCT-----GCTGGTTCCTCTTTTGGGG 447  
 ||| ||| : : : ||| : : :  
Db 629 oThrTrpSerSerThrSerAlaArgTrpArgThrArgSerProAlaSerArgTrpG 649  
Qy 446 CTCTCCATTACTGGGCTCCGCGACTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 387  
 ||| ||| : : : ||| : : :  
Db 649 ySer-----TrpMetProSerSerMet-MetLeuLeuSerSerT 662  
Qy 386 TGCGG-----TCCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 345

Db 662 hrThrTrpGlnAlaArgThrArgAlaAlaSerTrpSerProLeuGlyLeuAlaArgSerL 682  
QY 344 CTGTGGGGTCCAGGTCTCTGGACTCGATACCCCAACAGCACCAGCTTCGCC 285  
Db 682 euLeuProLeuAlaThrAlaSerProCysArgArgThrProThrGly----- 697  
QY 284 TGAGCCCTCCACTCTCCGGCCCTCCCTCCCTAGTTATGGCCGTGACGAGACAAGAA 225  
Db 698 -----SerGlyProThrTrpArgSerCysSerSerTrpGlyThrGluArgHisArgA 715  
QY 224 GG-----TCCATGGCGTGGCACCAGCATGAGAGGAGCAGCAGCTTCGAGG 180  
Db 715 snTrpArgGlnCysGlySerGlnGlySerAlaArgMetArgArg----- 729  
QY 179 GCCTTCTCACTCTGAGTCACGTACACACACACCCCTCCGGAAGCTGCTGGTCTCTCTCA 120  
Db 730 -----ThrArgAlaAlaSerTrp-----T 736  
QY 119 CTCCTCACTCACTTCAACTCCAGTTGGATCTGCCTGTGGACCTACCCG---CCGTGTC 63  
Db 736 hrSerThrTrpTrpGluAlaSerSerThrCysCysTrpTrpProTrpGlyTrpProCyst 756  
QY 62 TCAGTAGCGGAGAAAGATCCAGACCTCAGG-----ACCCGAGTCCCGGCTCAC 12  
Db 756 rpSerSerProGlySerThrTrpSerThrGlySerCysAlaThrArg-CysProThrHis 775  
QY 11 AGCTCC 6  
Db 776 ProSer 777  
RESULT 34  
ID AAY83167 standard; Protein: 87 AA.  
XX AC AAY83167;  
XX AC  
XX AC  
DT 24-JUL-2000 (first entry)  
DE PAGE1 polypeptide.  
XX KW PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;  
XX KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;  
KW cytotoxic T lymphocyte; immune response; antibody; drug delivery;  
KW immunconjugate.  
XX OS Homo sapiens.  
XX PN W0200012706-A1.  
XX PD 09-MAR-2000.  
XX PF 31-AUG-1999; 99WO-US20046.  
XX PR 01-SEP-1998; 98US-0098993.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX PI Pastan I, Brinkmann U, Vasmatazis G, Lee B;  
XX DR WPI; 2000-237869/20.  
XX DR  
XX DR  
PT Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T  
PT lymphocyte response and for raising antibodies which can be used to  
PT detect the presence of PAGE-4 in cell samples or body tissues  
XX  
XX Disclosure; Figure 1b; 63pp; English.  
XX  
CC PAGE-4 is a gene preferentially expressed in normal male and female  
CC reproductive tissues e.g. prostate, testis, fallopian tube, uterus  
CC and placenta, as well as in prostate cancer, testicular cancer and  
CC uterine cancer. This expression pattern makes it a target for  
CC diagnosis and for vaccine based therapy of such neoplasms.

CC An isolated PAGE-4 peptide which induces a cytotoxic T  
CC lymphocyte response when bound to a major histocompatibility complex  
CC (MHC) class I molecule or the isolated PAGE-4 protein can be used in  
CC immunogenic compositions to raise a cytotoxic T lymphocyte response  
CC against cells expressing PAGE-4 including cancer cells of the  
CC prostate, uterus and testis. The nucleic acids encoding PAGE-4 or  
CC PAGE-4 peptide fragments can also be used in these compositions.  
CC Antibodies against PAGE-4 and its peptide fragments can be used in  
CC detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell  
CC samples or body tissues. The presence of PAGE-4 in tissues which are  
CC not related to reproduction can be indicative of the spread of  
CC cancerous reproductive tissue. PAGE-4 can also be used to raise  
CC antibodies which are then used as the targeting group of  
CC immunoconjugates comprising toxins used in therapeutic applications.  
CC This has applications for drug delivery systems. This sequence is  
CC the PAGE1 polypeptide which shares sequence similarity with  
CC PAGE-4.  
XX  
SQ Sequence 87 AA;  
Alignment Scores:  
Pred. No.: 0.148 Length: 87  
Score: 100.00 Matches: 28  
Percent Similarity: 44.30% Conservative: 7  
Best Local Similarity: 35.44% Mismatches: 25  
Query Match: 7.25% Indels: 19  
DB: 21 Gaps: 1  
US-09-658-824-808 (1-781) x AAY83167 (1-87)  
QY 441 GGAGAGCCCCAAAAGAACACAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAG 500  
Db 28 GlyGluSerGlnGluGluProProThrAspAsnGlnGlyPro----- 42  
QY 501 ACAGAGAAGATCAGGATACAGCTGAGATCCCGTGGCGACATGGAAGGTGATCTGCAA 560  
Db 43 -----AspMetGluAlaPheGlnGln 49  
QY 561 GAGCTGCATCAGTCAAAACACCGGGATATAATCTGGTTTCCGCGCTCAAGGTGAA 620  
Db 50 GluLeuAspLeuGlyThrArgSerGluArgGlyAsp-GlySerAspVallysGluLy 69  
QY 621 GATAATACCTAAAGAGAACACTGTAAATGCCAGAGCAGGTGAAGAGCAACCA 675  
Db 69 sThrProAsnProLysHisAlaLysThrLysGluAlaGlyAspGlyGlnPro 87  
RESULT 35  
ID AAM95890 standard; Protein: 203 AA.  
XX AC AAM95890;  
XX AC  
DT 21-NOV-2001 (first entry)  
XX  
DE Human reproductive system related antigen SEQ ID NO: 4548.  
KW Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy.  
XX  
XX Homo sapiens.  
XX  
PN W0200155320-A2.  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01339.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-02321968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234597.  
PR 25-SEP-2000; 2000US-0234598.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0251989.  
PR 06-DEC-2000; 2000US-0251479.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX  
(HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465570/50.  
DR N-PSDB; AAL01860.  
XX

Isolated nucleic acid molecule encoding a reproductive system antigen  
is used in preventing, treating or ameliorating a medical condition -

Claim 11; SEQ ID NO 4548; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a  
number of human reproductive system related antigens. These can be used  
in the prevention and treatment of reproductive system disorders,  
CC

```
CC including cancer. The present sequence is a protein of the invention.
XX
SQ Sequence 203 AA;

Alignment Scores:
Pred. No.: 0.201 Length: 203
Score: 100.00 Matches: 48
Percent Similarity: 39.39% Conservative: 17
Best Local Similarity: 29.09% Mismatches: 53
Query Match: 7.25% Indels: 47
DB: 22 Gaps: 8

US-09-658-824-808 (1-781) x AAM95890 (1-203)
QY 128 AACACGAGGCTCCGAGGTTGTGTGTCAGTCACTGAGTCAAGCCCTCGAAG 187
DB 36 HSGlnAlaGlyAspGlyClyArg-----LeuArgLeuLeuArgAlaAlaLys 51
QY 188 TCGTCGTCCTCTCATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAA 247
DB 52 -----GlyAlaAspPro----- 55
QY 248 CTAGGAGGAGGCGGCGGAGGTGAGGGCTCAGCGGAAGCTGGGTCTCTTGGG 307
DB 56 ---GlyThrGlnArgIleArgGlyValHisGlyArgHisArg----- 68
QY 308 GGTATCCGAGTCCACAGACCTGGAACCCCGACAGAAAGATTCTGGACTCCCGACAGCGG 367
DB 69 -----ProGlyArgGlyThrGluSerGlu-----ProLysGly 79
QY 368 GACACGAGGAGCGGATGAGCGACACACACAAACACAGAACACACAGCCAGTCCCA 427
DB 80 GluSerPheAspSerGlyValSerSer-IleGlyThrGluProAspSerValGluG1 99
QY 428 GGAG-----CCGAGTAATGGAGCGCCCAAAAGAAAGAACACGAC 466
DB 99 nGlnPheGlyProGlyAlaAlaArgAspSerGlnAlaGluProThrGlnGluGlnAl 119
QY 467 ACCTGAAGTCGGGATCCTACACCTTGGCAGCAGACAGAAAGATGATACAGTCA 526
DB 119 aAlaGluAla---ProAlaGluGlyGlyProGlnThrAsnGlnLeuGluThrGlyAlaSe 138
QY 527 GATCCCGAGTGGC-----GACATGAAGGTGATCTGCAAGAGCTGCATCAGTCAACAC 580
DB 138 rSerProGluArgSerAsnGluValGluMetAspSerThrValIleThrValSerAsnSe 158
QY 581 CGGGGATAAATCT 593
DB 158 rSerAspLysSer 162

RESULT 36
AAM57046
ID AAM57046 standard; Protein; 411 AA.
XX
AC AAM57046;
XX
DT 20-AUG-1998 (first entry)
XX
DE Mouse apoptosis inducing receptor.
XX
KW Apoptosis inducing receptor; AIR protein; mouse; cell death regulator;
KW Type I transmembrane protein; tumour cell death; autoimmune disease;
KW therapy.
XX
OS Mus sp.
XX
PN W09814565-A1.
XX
PD 09-APR-1998.
XX
PF 03-OCT-1997; 97W0-US17876.
XX
PR 04-OCT-1996; 96US-0044456.
```

```
XX (IMMV ) IMMUNEX CORP.
PA
XX Perkins PA;
PI
XX
XX WPI; 1998-240077/21.
DR N-PSDB; AAV28701.
XX
XX DNA encoding apoptosis inducing receptor - which is Type I
PT transmembrane protein, useful for regulating cell death
XX
XX Claim 16; Page 33-35; 45pp; English.
XX
CC This sequence is the mouse apoptosis inducing receptor (AIR) of the
CC invention. AIR is a Type I transmembrane protein, soluble forms of which
CC can be used to regulate cell death in a therapeutic setting. Soluble AIR
CC can also be used in vitro to block apoptosis or AIR-expressing cells, or
CC to screen agonists or antagonists of AIR activity. The cytoplasmic domain
CC of AIR can be used to develop assays for inhibitors of AIR-induced cell
CC death, which is useful to regulate cell death in a therapeutic setting as
CC well as in vitro. Agonists of AIR activity can be used to kill tumour
CC cells that express AIR, or T cells expressing AIR in autoimmune diseases.
XX
SQ Sequence 411 AA;

Alignment Scores:
Pred. No.: 0.325 Length: 411
Score: 99.00 Matches: 63
Percent Similarity: 33.98% Conservative: 24
Best Local Similarity: 24.61% Mismatches: 70
Query Match: 6.94% Indels: 99
DB: 19 Gaps: 14

US-09-658-824-808 (1-781) x AAM57046 (1-411)
QY 677 TGTGTTCTCTTCCACCTGCTTCTGGCATTTTACAGTGTCTCTTTAGGTATTATCTTC 618
DB 135 CysGlyLysSerSerProPheSer-----CysValProCysGly-AlaThrTh 150
QY 617 ACCTTGAGCGCGGAACCCAAATCCA---GATTATCCCGCGTGT-----CA 573
DB 150 rProValHisGluAlaProThrProArgProCysLeuProGlyPheThrIleArgGlyAs 170
QY 572 CTGATGACAGCTTCTGCAGATCACCTTCCATGTCGCGCACTGGGATCTCAGCTGATCTCGT 513
DB 170 nAspCysThrSerCys-----ProThrGlyPheSerSerVal----- 182
QY 512 ATCTTCTTCTCTGCTGCCAGGTAGATCCGACTTTCAGCTGCTGTTCTTCTTT 453
DB 183 -----CysProLysAlaCysThrAlaValCysGlyTrpLysGlnMetPh 197
QY 452 TTGGGGCTCTCCATTACTGGGCTCTCTGGGACTGGCTGTGTCTGTGTTGTGTGTGT 393
DB 197 eTrpValGlnValLeuLeuGlyVal---AlaPheLeuPheGlyAlaIleLeuIleCysAl 216
QY 392 CGCTCATGCGCTCCCTCTCTGTCGTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTC 333
DB 216 a-----TyrCysArgTrpG1 221
QY 332 CAGGTGC-----TTCCTGGGACTCGGATACCCCAACAGC---ACCCAGCTTCGCTGA 282
DB 221 nProCysLysAlaValValThrAlaAspThrAlaGlyThrGluThrLeuAlaSerProG1 241
QY 281 GCCCTCCACCTCTCGGCCCTCTCTCTCCCTAGTTATGGCCGTGACGAGACAAGAAGT 222
DB 241 nThrAlaHisLeuSerAla----- 247
QY 221 CCATGGGCGTGGCACCAGCATGAGAGGACGACGACTTCGAGGGCCTTCTCACTCTGAGTC 162
DB 248 -----SerAspSe 250
QY 161 ACTGACCACACAAACCCCTCCGGAAGCCCTGCTGTTCTTCTTCTCACTCTCACTCACTCAA 102
DB 161 ----- 102
```

US-09-658-824-808 (1-781) x ABB60019 (1-773)	
Qy	221 GACCTCTTGTCTGTCACGCCCATACCTAGGAGGAGGAGCGCGAGAGTGAGGGG 280 
Db	658 AspLeuLeuThrGlnAsnArgPheAlaLeuArgLysThrGlyArgSerLeuGluIle 677 
Qy	281 CTCAGCGAAGCTGGGTGCTGTTGGGGTATCCGAGTCCCAAGCACTCTGAACCCCG 340 
Db	678 ProGluGluGlnValAlaProLysGlyAspLeuGlnLysGluGluLysProLysGluGlu 697 
Qy	341 ACAGAAGATTCTGACTCCCCAGACGGGACGAGGAGGACGGCATGAGCGACACACAC 400 
Db	698 GluGlnLysGluLysLeuProLysGluGluValGln-----LeuGluGluIleLys 714 
Qy	401 AAACACAGAACCCACACAGC-----CACTCCAGGAGCCCGAGTATGGAGAGCCCA 451 
Db	715 LysGlu-GluProGlnLysGluGluLeuGlnLysGluGluProGlnLysGluGluProGln 734 
Qy	452 AAAAGAAGAACCCAGCAGCTGAA-----AGTCGGGATCTACACCTGGGCGAGCAGACAGA 505 
Db	734 nLysGluGluProArgLysGluGluProGlnLysGluGluProGlnLysGluGluProGln 754 
Qy	506 AGAAGATCAGGATACAGCTGAGATCCAGTCCGCGACATGGAAGGTGATCTGCAAGAGCT 565 
Db	754 nLysGluGluProLysValGluThrPro-----GlnProLe 766 
Qy	566 GCATCAGTCA 575 
Db	766 uGluGlnSer 769 
RESULT 38	
ABB67402	
ID	ABB67402 standard; Protein; 821 AA.
AC	ABB67402;
XX	
DT	26-MAR-2002 (first entry)
XX	Drosophila melanogaster polypeptide SEQ ID NO 28998.
DE	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
XX	
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE ) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656860/75.
DR	N-PSDB; ABL11505.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PT	
XX	
PS	Disclosure; SEQ ID NO 28998; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA



CC sequences (AB01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 821 AA;

Alignment Scores:

Pred. No.: 0.415 Length: 821  
Score: 99.00 Matches: 34  
Percent Similarity: 42.74% Conservative: 19  
Best Local Similarity: 27.42% Mismatches: 54  
Query Match: 7.17% Indels: 17  
DB: 22 Gaps: 4

US-09-658-824-808 (1-781) x ABB67402 (1-821)

QY 221 GACCTTCTGCTCCTCAGCGCCATACTAGGAGGAAGGAGGCGGAGGAGTGGAGGG 280  
DB 706 ASPLeuLeuThrGlnAsnArgArgPheAlaLeuArgLysThrGlyArgSerLeuGluLe 725  
QY 281 CTCAGCGAAGCTGGGGTCTGCTGGGGGTATCCGAGTCCAGAACGACCTTGGAAACCCG 340  
DB 726 ProGluGluGlnValAlaProLysGlyAspLeuGlnLysGluGluLysProLysGluGlu 745  
QY 341 ACAGAGATCTGGACTCCCGACAGCGGACAGGAGGAGGCGGATGAGCACACAC 400  
DB 746 GluGlnGluLysGluLysProLysGluGluValGln-----LeuGluGluLeuLys 762  
QY 401 AAACAGACAGACACACAGC-----CAGTCCAGGAGCGCCAGTAAATGGAGGCCCA 451  
DB 763 LysGluGluProGlnLysGluGluLeuGlnLysGluGluProGlnLysGluGluProGln 782  
QY 452 AAAGAGAACACGAGCTGAA-----AGTCGGGATCTACACCTGGCGAGCAGACAGA 505  
DB 782 nLysGluGluProArgLysGluGluProGlnLysGluGluProGlnLysGluGluProGln 802  
QY 506 AGAAGATCAGGATACAGCTGAGATCCAGTCCGAGTCCGAGGAGTGTCTCAAGAGCT 565  
DB 802 nLysGluGluProLysValGluThrPro-----GlnProLeu 814  
QY 566 GCATCAGTCA 575  
DB 814 uGluGlnSer 817

RESULT 39

ABG29246  
ID ABG29246 standard; Protein; 494 AA.

XX AC ABG29246;

DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #29237.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.  
DR N-PSDB; AAS93433.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 20; SEQ ID No 59605; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 494 AA;

Alignment Scores:

Pred. No.: 0.39 Length: 494  
Score: 98.50 Matches: 66  
Percent Similarity: 35.00% Conservative: 32  
Best Local Similarity: 23.57% Mismatches: 74  
Query Match: 6.91% Indels: 109  
DB: 22 Gaps: 14

US-09-658-824-808 (1-781) x ABG29246 (1-494)

QY 712 AGCTTCGGTGTGTTGAGCTTGCT---TCATTTAACTGTGGTGTCTTCAC----- 662

DB 168 SerLeuArgCysLeuLeuValSerPheSerCysPheLeuCysValAlaLeuArgSer 18;

QY 661 -----CTGCTTCGGCATTTTACAGTCTTCCTCTT 632

DB 188 SerValPheLeuProPheCysLeuSerLeuPheCysAlaAlaPheArgAlaLeuProLeu 207

QY 631 TAGGTA-----TTATCTTCACCTTGACCGCCGAACCAATCCAGATTTA--- 587

DB 208 SerAlaProProSerPheValAlaCysProValLeuArgSerSerProProValCys 227

QY 586 -----TCCCGGTGTTGACTGATGAGCTCTTGAGATCAGCTCC 545

DB 228 ArgValSerArgSerSerValValLeuProProCysAlaCysAlaSerProLeu 247

QY 544 ATGTCGCGCAGCTGGGATCTC---AGCTGTATCTCTGCTCTGCTGCCAGG 489

DB 248 LeuPheAlaSerSerValCysArgSerCysValProLeuPhePheCysLeuSerLeu 267

QY 488 TGTAGGATCCCGACTTTC-----AGCTGCTGGTCTTCTTTTGGGGCTCCATT 438

DB 268 CysSerLeuProLeuPheLeuLeuCysAlaCys---PheProAlaLeuValSerLeu 286

QY 437 ACTGGGCTCCTGGGA-----CTGGCTGTCTGG-----TTCGTGTT 402

DB 287 ProArgLeuPheValPheProLeuLeuPheValTrpLeuPheValAla 306

QY 401 TGTGTGTGT----- 393  
 Db 307 CysValCysValAspCysLeuCysSerSerLeuHisSerProArgArgGlyAlaGln 326  
 QY 392 -----CGTCATCGCTCCCTCT----- 375  
 Db 327 TrpProArgThrCysGlnGluAsnLysAlaGlyAlaAlaArgTrpAsnAspGluThrAla 346  
 QY 374 -----CCTGGTCCCGTC--- 363  
 Db 347 HisLeuAlaAlaGluGlyArgAlaGluLysSerProGlnProProAlaProLeuSer 366  
 QY 362 -----TGGGAGTCCAGAAATCTTCTGTCGGGTTCACGGTCTT 324  
 Db 367 CysLeuLeuValAspLeuProSerGlyProGluGluAlaGluAspHisPheLeuAlaArg 386  
 QY 323 CTGGGACTGGATACCCCAACAGCACCCAGCTTCGCTGAGCCCTCCACTCCPCGCGC 264  
 Db 387 ProGlyLeuGlyTyrCysArgArgArg-ProArgSerArgAlaProProAlaAlaAlaPr 406  
 QY 263 CCTCCTCTCCCTAGTTATGCGCGTGACGAGACAAAGGTCATGCGGTCGCGCACCGC 204  
 Db 406 OCysPheProPro-----SerArgLeuHisValGlnGlyLeuAl 419  
 QY 203 ATGAGAGGACGACGACTTCGAGGGCTTCTCACTCTGAGTCACTGACACACAAAGCC 146  
 Db 419 a-----AlaArgSerArgProLeuThrSerAlaPro 429  
 RESULT 40  
 ID AAE02397  
 XX AAE02397 standard; Protein; 522 AA.  
 AC AAE02397;  
 DT 06-AUG-2001 (first entry)  
 XX  
 DE Canine retinitis pigmentosa GTPase regulator (RPGR) protein.  
 KW Dog; X-linked progressive retinal atrophy; XLPRA; genetic marker;  
 KW retinitis pigmentosa GTPase regulator; RPGR; Siberian Husky; Samoyed;  
 KW Miniature Schnauzer.  
 XX  
 OS Canis familiaris.  
 XX  
 PN WO200138578-A1.  
 XX  
 PD 31-MAY-2001.  
 XX  
 PF 21-NOV-2000; 2000WO-US31940.  
 XX  
 PR 24-NOV-1999; 99US-0167365.  
 XX  
 PA (CORR ) CORNELL RES FOUND INC.  
 XX  
 PI Aguirre GD, Acland GM, Zhang Q, Ray K, Zeiss CJ;  
 XX  
 DR WPI; 2001-367707/38.  
 DR N-PSDB; AAD07130.  
 XX  
 PT Identifying dogs with or carrying X-linked progressive retinal atrophy  
 PT by detecting retinitis pigmentosa GTPase regulator gene mutation,  
 PT useful when breeding Husky, Samoyed and Miniature Schnauzer -  
 XX  
 PS Claim 72; Page 31-32; 88pp; English.  
 XX  
 CC The invention relates to a method for identifying dogs which are  
 CC genetically normal, are carriers of, or are affected with X-linked  
 CC progressive retinal atrophy (XLPRA), by testing a biological sample with  
 CC genetic markers that co-segregate with a XLPRA gene locus. The invention  
 CC also relates to canine retinitis pigmentosa GTPase regulator (RPGR), RPGR  
 CC mutants and their corresponding nucleic acid molecules. The mutated RPGR  
 CC genes are responsible for the XLPRA in dogs. Methods are used to select  
 CC dogs for breeding so that dogs carrying the mutated locus are eliminated

CC from the breeding stock. The method particularly applies to Siberian  
 CC Husky, Samoyed and Miniature Schnauzer breeds, or any other breed where  
 CC the disease is X-linked. XLPRA1 type is identified in Siberian Huskies,  
 CC and Samoyeds, while XLPRA2 type is identified in Miniature Schnauzers.  
 CC The present sequence is Canine retinitis pigmentosa GTPase regulator  
 CC (RPGR) protein which is obtained from the normal open reading frame  
 CC (ORF) 15 of canine RPGR cDNA.  
 XX  
 SQ Sequence 522 AA;  
 Alignment Scores:  
 Pred. No.: 0.397 Length: 522  
 Score: 98.50 Matches: 52  
 Percent Similarity: 37.17% Conservative: 32  
 Best Local Similarity: 23.01% Mismatches: 112  
 Query Match: 7.14% Indels: 30  
 DB: 22 Gaps: 7  
 US-09-658-824-808 (1-781) x AAE02397 (1-522)  
 QY 69 GCGGTAGTCCACAGGAGATCCAACTGGAGTTGAAG----- 107  
 Db 224 AlaGlyGlyGluGluGlyGluAlaGluTrpGluGlyLysGlyValGluGluGlyGlu 243  
 QY 108 TGTGAGTGAGAGTGAAGAGGAAACAGCAGCGCTCCGAGGGTGTGTGTCAGTGACTCA 167  
 Db 244 GlyGluValGluGluGluGluGluGluValGluGluGlyGlyGlyGlyGlyGly 263  
 QY 168 GAG---TGAGAGGCGCTCGAAGTGTCTCTCTCATCGGTGCCAGCCCATGGACC 224  
 Db 264 GluValGluGluGlyGluAlaGluGluAspAlaGluGluGluGluArgGlu----- 281  
 QY 225 TTCTTGTCTCAGCGGCCATACTAGGAGGAAGAGG----- 263  
 Db 282 ---ValGluGluGluGlyGlyGlyGlyAsnLysArgGluTrpGluGluGluGlu 300  
 QY 264 -----GCCGAGGAGTGGAGGCGCTCAGCGAGCTCGAGTGGGTGCTGTG 305  
 Db 301 GluGlyGluGluGluGluGluArgGluAsnGluArgGluGluGluGlyGlyGluAla 320  
 QY 306 GGGGTATCCGAGTCCCAAGAACCCCTGGAACCCGACAGAGATTCTGACTCCCGACAG 365  
 Db 321 GlyGlyGluGluGluGluGluGluAlaGlyGlnGluGluGlyGlyGlyGluGlu 340  
 QY 366 GGGACACGAGAGGAGCGCATGAGCGACACACAAACAGACACAGACAGCCAGCCATCC 425  
 Db 341 GlyGluGluGluValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlu 360  
 QY 426 CAGGAGCCAGTAATGGAGAGCCCAAGAAAGAACAGCAGCTGAAAGTCGGATCCT 485  
 Db 361 GluGlu-----GlyGluGluGlyGlyGluGluGluGlyGlyGlyGlyGlyGlu 377  
 QY 486 ACACCTGGGACGACAGAGAAGATCAGGATACAGTGTAGATCCCGACGACATG 545  
 Db 378 GlyArgGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 397  
 QY 546 GAAGGTGATCTCAA---GAGCTGCATCAGTCAACACCGGGGATAAATCTGGATTGGG 602  
 Db 398 GluGlyAspLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 417  
 QY 603 TTCCGGCGTCAAGGTGAAGATAATACCTAAAGAGGAACACTCTAAATACCCAGAGCAGG 662  
 Db 418 ValArgGluArgGluGluLysAspThr-GluGluGlyGly---LysTyrGluGluThrGlu 436  
 QY 663 TCAAGAGCAACCAAA 678  
 Db 436 yAspGluGluSerGlu 441  
 Search completed: October 18, 2002, 09:44:34  
 Job time : 66 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on:            October 18, 2002, 09:39:56 ; Search time 50 Seconds  
                       (without alignments)  
                       5404.362 Million cell up

**Title:** US-09-658-824-808

Perfect score: 1380  
Sequence: 1 gcggcgagctatgacgg.....cttcacccccaaaaaaa 781

Scoring table: BLOSUM62

Scoring table.	0.00000	0.5
	Xgapop 10.0 ,	Xgapext 0.5
	Ygapop 10.0 ,	Ygapext 0.5
	Fgapop 6.0 ,	Fgapext 7.0
	Delop 6.0 ,	Delext 7.0

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 11244444

Minimum DB seq length: 0

Minimum DB seq	length: 0
Maximum DB seq	length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

```

COMMAND LINE PARAMETERS:
-MODEL=framet_n2p.model -DEV=xlpb
-CG/cg12/1/USTPO.spool/US09568824/runat_18102002_093947_9942/app_query.fasta_1.967
-DB=SPREMBL19 -QFMT=faston -SUFFIX=rspt -MINMATCH=0.1 -LOOPCPI=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09568824.@CGN1.1.88@runat_18102002_093947_9942 -NCPUP=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQRY=NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-VGAPOP=10 -VGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database :

```

1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.rvirus:*
16: sp.bacteriap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	DB Length	ID	Description
1	425	30.8	81	4 Q969J6	Q969J6 homo sapien

2	233	16.9	111	4	Q96GT9	Q96GT9	homo sapien
3	109	7.9	146	4	Q9BS57	Q9BS57	homo sapien
4	109	7.9	888	2	Q93GW8	Q93GW8	streptomyce
5	108	7.8	1003	12	Q91LX9	Q91LX9	kaposi's sa
6	106.5	7.7	976	12	Q9DUN0	Q9DUN0	kaposi's sa
7	106	7.7	398	4	Q96D06	Q96D06	homo sapien
8	106	7.7	1045	10	Q94HT6	Q94HT6	oryza sativ
9	105.5	7.6	528	4	Q9P0W9	Q9P0W9	homo sapien
10	104.5	7.6	355	2	Q918W2	Q918W2	mycoplasma
11	104.5	7.6	384	2	Q918V9	Q918V9	mycoplasma
12	104.5	7.6	538	11	Q9RT15	Q9RT15	mus musculus
13	104	7.5	319	16	Q9RV01	Q9RV01	deinococcus
14	103.5	7.5	498	11	Q9CSL7	Q9CSL7	mus musculus
15	103	7.5	246	11	Q63134	Q63134	rattus norv
16	103	7.2	413	11	Q99MM1	Q99MM1	mus musculus
17	103	7.5	643	5	Q9U229	Q9U229	caenorhabdi
18	102.5	7.4	1036	12	Q9DUM3	Q9DUM3	kaposi's sa
19	102.5	7.4	1162	12	Q9B148	Q9B148	kaposi's sa
20	102.5	7.2	2971	4	Q9Y5L9	Q9Y5L9	homo sapien
21	102	7.2	350	8	Q33572	Q33572	trypanosoma
22	102	7.2	1285	4	Q96JL1	Q96JL1	homo sapien
23	100.5	7.3	1054	10	Q9SN08	Q9SN08	oryza sativ
24	100	7.2	480	5	Q27033	Q27033	theileria p
25	99.5	7.2	181	10	Q94E85	Q94E85	oryza sativ
26	99.5	7.0	325	4	Q15627	Q15627	homo sapien
27	99	7.2	116	4	Q9UEU5	Q9UEU5	homo sapien
28	99	7.2	630	11	Q99JV6	Q99JV6	mus musculus
29	99	7.2	773	5	Q76871	Q76871	drosophila
30	99	7.2	773	5	Q94AV0	Q94AV0	drosophila
31	99	7.2	821	5	Q917X0	Q917X0	drosophila
32	98.5	7.1	405	10	Q41516	Q41516	triticum ae
33	98.5	7.1	405	10	Q9SYR0	Q9SYR0	triticum ae
34	98.5	7.1	847	6	Q95KP7	Q95KP7	canis famli
35	98.5	7.1	1012	10	Q949B8	Q949B8	oryza sativ
36	98.5	7.1	1422	6	Q95KU4	Q95KU4	canis famli
37	98	7.1	299	4	Q13059	Q13059	homo sapien
38	98	7.1	1089	12	Q40947	Q40947	kaposi's sa
39	98	7.1	1245	4	Q43636	Q43636	homo sapien
40	98	7.1	1251	4	Q9UMG2	Q9UMG2	homo sapien
41	98	6.9	1668	4	Q15026	Q15026	human herpe
42	97.5	7.1	296	12	Q69118	Q69118	human herpe
43	97	7.0	345	5	Q9VJ23	Q9VJ23	drosophila
44	97	6.8	417	5	Q9V6Q3	Q9V6Q3	drosophila
45	97	7.0	457	11	Q8B344	Q8B344	rattus norv

## ALIGNMENTS

RESULT 1				
Q969J6				
ID	Q969J6	PRELIMINARY;	PRT;	81 AA.
AC	Q969J6;			
DT	01-DEC-2001	(TREMblrel. 19, Created)		
DT	01-DEC-2001	(TREMblrel. 19, Last sequence update)		
DT	01-DEC-2001	(TREMblrel. 19, Last annotation update)		
DE	CANCER/TESTIS-ASSOCIATED PROTEIN XAGE-1B (HYPOTHETICAL 9.1 KDA DE PROTEIN) (9 KDA CANCER/TESTIS ASSOCIATED PROTEIN).			

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

RP	SEQUENCE FROM N.A.
RP	TISSUE=MELANOMA METASTASIS;
RC	Zendman A.J.W., van Kraats A.A., Weidle U.H., Ruiter D.J.
RC	*XAGS-1b, a novel cancer/testis-associated gene, induced
RA	metastases.";
RT	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RT	[2]
RL	SEQUENCE FROM N.A.
RP	TISSUE=LUNG CARCINOMA, AND LARGE CELL UNDIFFERENTIATED;
RC	



```

RESULT 4
Q93GW8      PRELIMINARY;      PRT;      888 AA.
AC   Q93GW8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE LARGE SECRETED PROTEIN.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horioka H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   *Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.*;
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070957; BAB69423.1; -.
SQ   SEQUENCE 888 AA; 93501 MW; 061EC2522217C15F CRC64;

Alignment Scores:
Pred. No.:      0.0618      Length:      888
Score:          109.00      Matches:    45
Percent Similarity: 35.95%      Conservative: 10
Best Local Similarity: 29.41%      Mismatches: 64
Query Match:      7.90%      Indels:     35
DB:              2          Gaps:         7

US-09-658-824-808 (1-781) x Q93GW8 (1-888)
QY 238 ACGCCATACTAGGAGGAGGCGCGAGAGTGGAGGGCTCAGGGGAGCTGGG 297
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 684 ThrArgLeuProAspHisSerGlyPheProGluTyrglyGlyAla---GlyLeuLeuPro 702
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 298 TGCTGTGGGGTATCCGAGT-----CCCAAGACACCTGGAACCCCGACAGA 345
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 703 AlaAlaSerAspHisProAlaProAlaProGlyProAsnAlaSerTrpAlaProArg 722
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 346 AGATTCTGGACTCCCGAG-----ACGGGACACGAG----- 375
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 723 GluHisProAlaProAlaProGlnGlyHisArgThrGluProGlyGlyLeuGlyProGlnAlaGly 742
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 376 -----GAGGACGCGCATGCGCACACACACAAAC-----ACAGAA 410
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 743 GlnProAlaHisProGlyAlaHisAlaGlyHisHisGlySerArgAlaGlyPheThrGlu 762
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 411 CCACACACCCAGTCCAGGAGCCCGAGTAAATGGAGAGCCCGCCCAAGAGAACGACGCT 470
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 763 ProGlnSerGlyPheThrGluProArgSerGlyProProSerArgHisGluProGlyAla 782
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 471 GAAAGTCGGGATCCTACAGCTGGGAGCAGACAGAGAAGATCAGGATACAGCTGACATC 530
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 783 GlySerGlyGlyProAlaProAlaArgGlu-----GluAla 794
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 531 CCAGTGCAGCATGGAAGGTGATCTG-----CAAGAGCTGCATCATGCTAAACACC 581
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 795 ProAlaArgSerAlaAlaArgProLeuThrThrProGlnValLeuProGlnArgThrArg 814
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 582 GGGGATAAATCTGGATTTGGGTTCCGGCTCAAGGTGAA 620
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 815 GlyAlaSerLeuAlaGlnGlnLeuArgArgGluAlaGlu 827
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111

RESULT 5
Q91LX9      PRELIMINARY;      PRT;      1003 AA.
AC   Q91LX9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE LARGE SECRETED PROTEIN.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horioka H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   *Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.*;
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070957; BAB69423.1; -.
SQ   SEQUENCE 1003 AA; 115517 MW; C20C43308B01A0A3 CRC64;

Alignment Scores:
Pred. No.:      0.0786      Length:      1003
Score:          108.00      Matches:    28
Percent Similarity: 40.57%      Conservative: 15
Best Local Similarity: 26.42%      Mismatches: 41
Query Match:      7.83%      Indels:     22
DB:              2          Gaps:         2

US-09-658-824-808 (1-781) x Q91LX9 (1-1003)
QY 261 AGGCGGAGGAGTGGAGGGCTCAGCGAGCTGGGTGGGTATCCGAGTCC 320
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 409 LysLysGluAspGluGluAspGlyGlyAspGlyAsnLysThrLeuSerIleGlnSer 428
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 321 CAGAAGCACCTGGAACCCCGACAGAGATCTCGACTCCCGACGCGGACGAGAGGG 380
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 429 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 437
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 381 ACGGCATGAGCGACACACACACAAACACAGAACACACAGCCAGCCAGCCAGTAAT 440
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 438 -----GluPro-----GlnGlnGlnGluProGlnGln 446
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 441 GGAGAGCCCCAAAAGAAAGAACACGAGCTGAAAGTCGGGATCCTACACCTGGGACGAG 500
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 447 GlnGluProGlnGlnGlnGluProLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 466
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 501 ACAGAAGACATCAGATACAGCTGAGATCCCGACGATGCGCGCATGGAAGGTGATCTGCAA 560
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 467 GluProGlnGlnGlnGluProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 486
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 561 GAGCTGCATCAGTCAAAAC 578
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111
QY 487 GluProGlnGlnGlnAsp 492
Db 111 : : : : : 111 111 111 111 111 111 111 111 111 111

RESULT 6
Q9DUNO      PRELIMINARY;      PRT;      976 AA.
AC   Q9DUNO;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ORF73.
OS   Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammapherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   STRAIN=PK-1;
RX   MEDLINE=59445611; PubMed=10515805;
RA   Gao S.J., Zhang Y.J., Deng J.H., Rabkin C.S., Flore O., Jensen H.B.;
RT   *Molecular polymorphism of Kaposi's sarcoma-associated herpesvirus
RT   (Human herpesvirus 8) latent nuclear antigen: evidence for a large
RT   repertoire of viral genotypes and dual infection with different viral
RT   genotypes.*;
RL   J. Infect. Dis. 180:1466-1476(1999).

```

Accession	Score	Length	Matches	Conservative	Mismatches	Indels	Gaps
Q96D06	106.50	976	976	28	42	23	2
Q96D06	39.25%	976	976	28	42	23	2
Q96D06	26.17%	976	976	28	42	23	2
Q96D06	7.72%	976	976	28	42	23	2
Q96D06	12	976	976	28	42	23	2
Q96D06	0.109	976	976	28	42	23	2
Q96D06	106.50	976	976	28	42	23	2
Q96D06	39.25%	976	976	28	42	23	2
Q96D06	26.17%	976	976	28	42	23	2
Q96D06	7.72%	976	976	28	42	23	2
Q96D06	12	976	976	28	42	23	2
Q96D06	0.109	976	976	28	42	23	2
Q96D06	106.50	976	976	28	42	23	2
Q96D06	39.25%	976	976	28	42	23	2
Q96D06	26.17%	976	976	28	42	23	2
Q96D06	7.72%	976	976	28	42	23	2
Q96D06	12	976	976	28	42	23	2
Q96D06	0.109	976	976	28	42	23	2
Q96D06	106.50	976	976	28	42	23	2
Q96D06	39.25%	976	976	28	42	23	2
Q96D06	26.17%	976	976	28	42	23	2
Q96D06	7.72%	976	976	28	42	23	2
Q96D06	12	976	976	28	42	23	2
Q96D06	0.109	976	976	28	42	23	2
Q96D06	106.50	976	976	28	42	23	2
Q96D06	39.25%	976	976	28	42	23	2
Q96D06	26.17%	976	976	28	42	23	2
Q96D06	7.72%	976	976	28	42	23	2
Q96D06	12	976	976	28	42	23	2
Q96D06	0.109	976	976	28	42	23	2
Q96D06	106.50	976	976	28	42	23	2
Q96D06	39.25%	976	976	28	42	23	2
Q96D06	26.17%	976	976	28	42	23	2
Q96D06	7.72%	976	976	28	42	23	2
Q96D06	12	976	976	28	42	23	2
Q96D06	0.109	976	976	28	42	23	2
Q96D06	106.50	976	976	28	42	23	2
Q96D06	39.25%	976	976	28	42	23	2
Q96D06	26.17%	976	976	28	42	23	2
Q96D06	7.72%	976	976	28	42	23	2
Q96D06	12	976	976	28	42	23	2
Q96D06	0.109	976	976	28	42	23	2
Q96D06	106.50	976	976	28	42	23	2
Q96D06	39.25%	976	976	28	42	23	2
Q96D06	26.17%	976	976	28	42	23	2
Q96D06	7.72%	976	976	28	42	23	2
Q96D06	12	976	976	28	42	23	2
Q96D06	0.109	976	976	28	42	23	2
Q96D06	106.50	976	976	28	42	23	2
Q96D06	39.25%	976	976	28	42	23	2
Q96D06	26.17%	976	976	28	42	23	2
Q9							

Score:	106.00	Matches:	52		
Percent Similarity:	37.74%	Conservative:	28		
Best Local Similarity:	24.53%	Mismatches:	76		
Query Match:	7.66%	Indels:	56		
DB:	4	Gaps:	8		
US-09-658-824-808 (1-781) x Q96D06 (1-398)					
QY	52	TCCGCTACTGAGACACGGCGGTAGGTCACAGCAGATCCAACTGGGAGTTGAAGTGTG	1111		
			:::		
Db	101	SerValThrGluLeu--GlnThrHisProGluLeuAspThrAspGlyAspGlyAlaLeu	119		
QY	112	ACTGAGAGTGA-----GAGGAACACGACGCGTTC	141		
Db	120	SerGluAlaGluAlaGlnAlaLeuSerGlyAspThrGlnThrAspAlaThrSerPhe	139		
QY	142	CGAGGGTGTGTGTCAGTGACTGACAGAGGAGGCGCTCAAGTCGTCGCTCCCTCTC	201		
Db	140	TyrAspArgValTrpAlaAla-IleArgAspLysTyrArgSerGlu	154		
QY	202	ATGCGGTGCCACGCCATGGACCTTCTGTCTCGTCACGGCCAACTAGGGAGG---AA	258		
Db	155	-----AlaLeuProThrAspLeuProThrProSerAlaProAspLeuThrGluProLy	172		
QY	259	GGAGGGCGGAGGAGTGGAGGGGCTCAGCGCAAGCTGGGTGCTTGGGGGTATCCGAGT	318		
Db	172	sGluGluGln-----	175		
QY	319	CCCAGAAGCACTGGAACCCCGCAGAGAAGATCTTGACTCCCCAGACGGACACGAGAG	378		
Db	176	-ProValProSerSerProThrGluGlu-----GluGluGluGluGlu	192		
QY	379	GGAGCGCATGAGCAGACACACAACACAGACACACAGCCAGTCCCAGGAGCCAGTA	438		
Db	192	uGluGluAlaGluGluGluGlu-GluGluAspSerGluGluAlaProProProL	212		
QY	439	ATGGAGAGCCCAAGAGAACACAGCAGCTGAAGTCGGGATCTACACCTGGGCAGC	498		
Db	212	eUserProGlnProAlaSerProAlaGluGluAspLysMetProTyrAspGluG	232		
QY	499	ACACAGAAGAAGATCAGGATACA-----GCTGAGA	528		
Db	232	InThrGlnAlaPheIleAspAlaAlaGlnGluAlaArgAsnLysPheGluGluAlaG	252		
QY	529	TCCAGTTCGGGACATGGAAGGTGATCTCGAAGAGTGCATCAGTCAACACCGGGGATA	588		
Db	252	rgSerLeuLysAspMetGluSerIleArgAsnLeuGluGln-----G	267		
QY	589	AATCTGATTGGGTCCGCGCTCAAGGTGAA	620		
Db	267	IuIleSerPheAspPheGlyProAsnGlyGlu	277		
RESULT 8					
Q94HT6	PRELIMINARY;				
ID	Q94HT6	PRT; 1045 AA.			
AC	Q94HT6;				
DT	01-DEC-2001 (TREMBLrel. 19, Created)				
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
OS	HYPOTHETICAL PROTEIN.				
DE	Oryza sativa (Rice).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;				
OC	Ehrhartodiaceae; Oryzeae; Oryza.				
OX	NCBI_TaxID=4530;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-NIPPONBARE.				
RA	Du H., Minx P., Abbott A., Doebber A., de la Bastide M., Spiegel L.,				
RA	Nascimento L., Preston R., Kirchoff K., King L., Vill M.D., Baker J.,				
RA	Zutavern T., Santos L., Bell M., Miller B., Kuit K., Rodriguez S.,				
RA	Cunnius D.M., Ballja V., Shah R., Bahret A., O'Shaughnessy A.,				
RA	Palmer L., Yang C., Dedhia N., McComble W.R.,				

\*Genomic Sequence for *Oryza sativa*, Nipponbare strain, clone OSJNB0036B06, from Chromosome 10, complete sequence.";  
RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC051624; AAK92563.1; -  
SQ SEQUENCE 1045 AA: B834A297B81FE6F CRC64:

RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.
RA	Marcell A., Sevigny G., Thomas D.Y.;	RA	Marcell A., Sevigny G., Thomas D.Y.;
RT	*Expression of Enzymatically Active Isoforms of Human Glucosidase II	RT	*Expression of Enzymatically Active Isoforms of Human Glucosidase II
RL	In Insect Cells.*;	RL	In Insect Cells.*;
RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.	RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL: AF144075; AAF66686.1; ..	DR	EMBL: AF144075; AAF66686.1; ..
DR	InterPro: IPR002048; EF-hand.	DR	InterPro: IPR002048; EF-hand.
DR	InterPro: IPR000886; ER_target.	DR	InterPro: IPR000886; ER_target.
DR	InterPro: IPR002172; LDL_recept_A.	DR	InterPro: IPR002172; LDL_recept_A.
DR	SMART: SM00192; LDLa; 1.	DR	SMART: SM00192; LDLa; 1.
DR	PROSITE: PS00018; EF_HAND; UNKNOWN_1.	DR	PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR	PROSITE: PS00014; ER_TARGET; UNKNOWN_1.	DR	PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
SEQ	SEQUENCE 528 AA; 59425 MW; 8DAD9776037E878E CRC64;	SEQ	SEQUENCE 528 AA; 59425 MW; 8DAD9776037E878E CRC64;
Alignment Scores:			
Pred. No.:	0.124	Length:	528
Score:	105.50	Matches:	52
Percent Similarity:	37.74%	Conservative:	78
Best Local Similarity:	24.53%	Mismatches:	27
Query Match:	7.64%	Indels:	55
DB:	4	Gaps:	8
US-09-658-824-808 (1-781) x 09P0W9 (1-528)			
Qy	52 TCCGCTACTGACACACGGCGGTAGTCCACGAGCAGATCCAACTGGGAGTTGAAGTGTG	111	52 TCCGCTACTGACACACGGCGGTAGTCCACGAGCAGATCCAACTGGGAGTTGAAGTGTG
Db	230 SerValThrGluLeu---GlnThrHisProGluLeuAspThrAspGlyAspGlyAlaLeu	248	230 SerValThrGluLeu---GlnThrHisProGluLeuAspThrAspGlyAspGlyAlaLeu
Qy	112 ATGTCAGAGTGAA-----	141	112 ATGTCAGAGTGAA-----
Db	249 SerGluAlaGlnAlaLeuLeuSerGlyAspThrGlnThrAspAlaThrSerPhe	268	249 SerGluAlaGlnAlaLeuLeuSerGlyAspThrGlnThrAspAlaThrSerPhe
Qy	142 CGGAGGCTTGTTGGTTCAGTGACTCAGAGTGAGAGGCCCTCGAAGTCTGCTCCCTCTC	201	142 CGGAGGCTTGTTGGTTCAGTGACTCAGAGTGAGAGGCCCTCGAAGTCTGCTCCCTCTC
Db	269 TyrAspArgValIrrAlaAla--IleArgAspLysTyrArgSerGlu-----	283	269 TyrAspArgValIrrAlaAla--IleArgAspLysTyrArgSerGlu-----
Qy	202 ATGCGGTGCCACGCCATGACCTTCTGTCTCGACGCCATACTAGGAGG---AA	258	202 ATGCGGTGCCACGCCATGACCTTCTGTCTCGACGCCATACTAGGAGG---AA
Db	284 -----AlaLeuProThrAspLeuProAlaProSerAlaProAspLeuThrGluProly	301	284 -----AlaLeuProThrAspLeuProAlaProSerAlaProAspLeuThrGluProly
Qy	259 GGAGGCGCGAGGAGTGAGGGGCTCAGCGCAAGCTGGGTGCTGTGGGGGTATCCGAGT	318	259 GGAGGCGCGAGGAGTGAGGGGCTCAGCGCAAGCTGGGTGCTGTGGGGGTATCCGAGT
Db	301 sGluGluGln-----	304	301 sGluGluGln-----
Qy	319 CCCAGAGACACTGGAAACCCGACAGAGAAGATTCTGACTCCCGACAGCGGACACGAGAG	378	319 CCCAGAGACACTGGAAACCCGACAGAGAAGATTCTGACTCCCGACAGCGGACACGAGAG
Db	305 -ProProValProSerSerProThrGluGluGlu-----GluGluGluGluGluGluGlu	322	305 -ProProValProSerSerProThrGluGluGlu-----GluGluGluGluGluGluGlu
Qy	379 GGACGGCATGACGCACACACACAAACACACACACACAGCCAGTCCCAGGAGCCCACTA	438	379 GGACGGCATGACGCACACACACAAACACACACACACAGCCAGTCCCAGGAGCCCACTA
Db	322 uGluGluAlaGluGluGluGluGluGluGluGluAspSerGluGluAlaProProProL	342	322 uGluGluAlaGluGluGluGluGluGluGluGluAspSerGluGluAlaProProProL
Qy	439 ATGGAGAGCCCCAAAAGAAAGAACACGACGACTGAAAGTCCGGGATCCTACACCTGGCAGC	498	439 ATGGAGAGCCCCAAAAGAAAGAACACGACGACTGAAAGTCCGGGATCCTACACCTGGCAGC
Db	342 euSerProGlnProAlaSerProAlaGluGluAspLysMetProTyrAspGluG	362	342 euSerProGlnProAlaSerProAlaGluGluAspLysMetProTyrAspGluG
Qy	499 AGACAGAAGAAGATCAGGATACA-----	528	499 AGACAGAAGAAGATCAGGATACA-----
Db	362 lnrThrGlnAlaPheIleAspAlaAlaGlnGluAlaArgAsnLysPheGluGluAlaGlu	382	362 lnrThrGlnAlaPheIleAspAlaAlaGlnGluAlaArgAsnLysPheGluGluAlaGlu
Qy	529 TCCGAGTCGCGGACATGGAAGGTGATCTGCAAGAGCTGCATCATCAACACCGGGGATA	588	529 TCCGAGTCGCGGACATGGAAGGTGATCTGCAAGAGCTGCATCATCAACACCGGGGATA
Db	382 rgSerLeuLysAspMetGluGluSerIleArgAsnLeuGluGln-----	397	382 rgSerLeuLysAspMetGluGluSerIleArgAsnLeuGluGln-----
Qy	589 AATCTGATTTGGGTTCCGGCGTCAAGGTGAA	620	589 AATCTGATTTGGGTTCCGGCGTCAAGGTGAA
Db	397 luIleSerPheAspPheGlyProAsnGlyGlu	407	397 luIleSerPheAspPheGlyProAsnGlyGlu
RESULT 10			
ID	Q9L8W2	PRELIMINARY;	PRT; 355 AA.
AC	Q9L8W2;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		

Db	263	snThrGlyAsnLysThrThrSerCluSerAsnSerGlySerSerThrGlySerGluAlac	283
Qy	637	GAACACTGTAATCGACGAAGCAGGTGAGACACACCACTTTAAATGAGACAAG	695
Db	283	lyThrThrAsnThrGlySerGlySerAsnSerGluSerGlyMetAsnSerGluLys	302
RESULT 11			
Q9L8V9			

ID Q9L8V9 PRELIMINARY: PRT: 394 AA  
 100 amino acids 13 treated  
 11 sequence update)  
 12 first initiation update)  
 13 first initiation update)  
 14 first initiation update)  
 15 first initiation update)  
 16 first initiation update)  
 17 first initiation update)  
 18 first initiation update)  
 19 first initiation update)  
 20 first initiation update)  
 21 first initiation update)  
 22 first initiation update)  
 23 first initiation update)  
 24 first initiation update)  
 25 first initiation update)  
 26 first initiation update)  
 27 first initiation update)  
 28 first initiation update)  
 29 first initiation update)  
 30 first initiation update)  
 31 first initiation update)  
 32 first initiation update)  
 33 first initiation update)  
 34 first initiation update)  
 35 first initiation update)  
 36 first initiation update)  
 37 first initiation update)  
 38 first initiation update)  
 39 first initiation update)  
 40 first initiation update)  
 41 first initiation update)  
 42 first initiation update)  
 43 first initiation update)  
 44 first initiation update)  
 45 first initiation update)  
 46 first initiation update)  
 47 first initiation update)  
 48 first initiation update)  
 49 first initiation update)  
 50 first initiation update)

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2100;  
443

RP SEQUENCE FROM N.A.  
RC STRAIN-SK76;  
RX MEDLINE=20138163; PubMed=10671459;  
RA Citti C., Watson-McKown R., Droesse M., Wise K.S.;  
RT "Gene Families Encoding Phase- and Size-Variable Surface Lipoproteins  
of Mycoplasma hyorhinis";  
EL J. Bacteriol. 182:1356-1363(2000).  
DR EMBL; AF193880; AAF36549.1; -.

KW	Signal; Lipoprotein.	1	29	POTENTIAL.
FT	SIGNAL			

FT	CHAIN	30	384	87-KDA SURFACE LIPOPROTEIN.
SQ	SEQUENCE	384 AA;	38226 MW;	04A1F0A98E37B79C CRC64;

Alignment Scores:

Length:	384
Pred. No.:	0.148
Score:	104.50
Matches:	63

score.	104.50	Matches:	03
Percent Similarity:	39.83%	Conservative:	33

Best Local Similarity:	26.14%	Mismatches:	112
Query Match:	7.57%	Indels:	33

DB: 2 Gaps: 9

US-09-658-824-808 (1-781) x Q9L8V9 (1-384)

QY 16 GCCGGCGACTCGGGTCCCTGAGGTCTGGATTCTTTCTCCGCTACTGAGACACGGCGGTA 75

Db 111 SerGlySerAsnSerGluSerGlyMetAsnSerGluLysThrGluAsnThrGlnGlnSer 130

QY 76 GGTCCACAGGCAGATCCAACTGGGACTTGAAGTGTGAGTGAGAGTGAACAGCAA----- 129

db 131 GluAlaproGlyThrAsnThrGlyVAsnIysThrThrSerGlySerAsnSerGlySerSer 150

[illegible][illegible]

151 ThrGlySerGlnAlaGlyThrThrThrAsnThrGlySerGlySerAsnSerGluSerGly 170

Qy 181 CTC-----GAAGTCGTCGTCCTCTCATCGCGTGCCACGCCCATGGACCTTCTTGTCCTC 234

Db 171 MetAsnSerGluLysThrGluAsnThrGlnGlnSerGluAlaProGlyThrLys----- 188

QY 235 GTCACGGCCATAACTAGGGAGGAAGAGGGCCGAGGA-----GTGGAGGGGCTCAGGCG 288

Db 189 ---ThrGluAsnThrGlnGlnSerGluAlaPro-GlyThrLysThrGluAsnThrGlnG1 207

Qy 289 AAGCTGGGGTGCTGTTGGGGTATCCGAGTCCCAGAAGCACCTGGAAACCCCGACAGAAGA 348

Db 207 nSer-----GluAlaProGlyThrAsnThrGlyAs 217 ::  
: ||| ||||| ||||| |||

0V 349 T-----TCTGGACATCCCCACACCCGCCACCCGCCGCGATCAGCCGACACACA 300

[illegible]

21/ nlyslum m m ser -Gluse r as use r gylse r se r m r gylse r gln a r gylse r m r t 23/

400 CAATACAGAAACCCACACAGCCAGTCCCGAGGCCCGAGTATGGAGAGCCCCCAAAAG 459



```

Db 237 hrAsnThrGlySerGlySerAsnSerGluSerGlyMetAsnSerGluLysThrGluAsnT 257
QY 460 AACACAGACAGTGAAGTCCGGATCTACACCTGGGAGCAGACAGAGAGATCAGGATA 519
Db 257 hrGlnGlnSerGluAla-----ProGlyThrLysThrGluAsnThrGlnGln- 272
QY 520 CACTGAGATCCAGTCCGCGACATGGAAGTCTATCGCAA---GAGTCATCATCAACAA 576
Db 273 --SerGluAlaProGlyThrLysThrGluAsnThrGlnGlnSerGluAlaLeuGlyThrA 292
QY 577 ACACCGGGGATAATCTGGATTTCGGGCTCCGGCGCTCAAGGTGAAGATAATCAAGAG 636
Db 292 snThrGlyAsnLysThrSerGluSerAsnSerGlySerSerThrGlySerGlnAlaG 312
QY 637 GACACTGTAAATGCCAGACAGCAGGTGAAGACCAACCAACCAATTTAAATGAAGACAAG 695
Db 312 lyThrThrAsnThrGlySerGlySerAsnSerGluSerGlyMetAsnSerGluLys 331
RESULT 12
Q9ET15 PRELIMINARY; PRT; 538 AA.
AC Q9ET15;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DE RETINITIS PIGMENTOSA GTPASE REGULATOR (FRAGMENT).
GN RPR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Vervoort R., Lennon A., Bird A.C., Tulloch B., Axton R., Miano M.G.,
RA Meindl A., Meitinger T., Ciccodicola A., Wright A.F.,
RT "Mutational hot spot within a new RPR exon in x-linked retinitis
RT pigmentosa.";
RL Nat. Genet. 0:0-0(2000).
DR EMBL; AF286473; AAG00552.1; -.
FT NON_TER 1
SQ SEQUENCE 538 AA; 60333 MW; 549F3E092943FDE0 CRC64;

Alignment Scores:
Pred. No.: 0.156 Length: 538
Score: 104.50 Matches: 59
Percent Similarity: 38.60% Conservative: 29
Best Local Similarity: 25.88% Mismatches: 90
Query Match: 7.57% Indels: 51
DB: 11 Gaps: 10

US-09-658-824-808 (1-781) x Q9ET15 (1-538)
QY 109 GTGAGTGAGAGTGAAGAGGAACAGCAGGCTCCGGAGG-----GTTGTGTGTCAGTG 162
Db 1 IleProGluGlnGluGlyProGluAspSerGluGlyValValGluGlnVal 20
QY 163 ACTCAGAGTGAAGGCC---CTCGAAGTGTGTCCTCTCATGGTCCCGCCAGGCCAT 219
Db 21 ValGlnAlaGlnLysGluAsnLeuGluPheGluGlyAspArgLysGluAlaLysAlaGlu 40
QY 220 GGACCTCTTGTCTGCTCAGC-----GCCATACTAGGAGGAGGAGGG 264
Db 41 AlaProSerAspValIleThrGluLysGluValSerGluGluGluSerGlyGly 60
QY 265 CCAGAGAGTGG-AGGGCTCAGCGAAGCTGGGGTGTGTTG----- 305
Db 61 GluArgGluAspArgSerGluGlyAspGlyAspGlnIleCysGluLysValSerLeuGlu 80
QY 306 -----GGGTATCCGAGTCCAGAGCAGCAGACCTCGGAACCCGA 341
Db 81 ThrGluHisLeuGlnArgAlaGlnGlyLysGlnGluArgLysLysGlyLysAspLysArg 100
QY 342 CAGAAG---ATTCTGGACTCCCCAGAC-----GGGACCAGGAGGAGC 383

```

```

Db 101 AlaArgCysIleLeuAspMetLysGluArgGluGluAspLysGlyTrpGluLysGlySer 120
QY 384 GCATGAGCGCACACACAAACACAGAACACACAGCAGTC----- 425
Db 121 GluGlyGlyAspLysMetLysArgAspGluGlyAsnGlnGluLysArgLysLysGluMet 140
QY 426 CAGGAGCCCGAGTAAATGAGAGCCCCAAAAGAAAGAACACAGCTGAAGTCGGGATCCT 485
Db 141 GluLysArgAspAlaGlyAspGluArgSerGluGluGluGluGluGluPro 160
QY 486 ACACCTGGGCGAGCAGACA-----GAAGAAGATCAGGATACAGCTGAGATCCCCAGTG 536
Db 161 GluGluGlyGlyLysGluGluGlyGlyGluGluGluGlyThrSerGluAspGlnSer 180
QY 537 CGGACATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAACACCGGGGATAAATCTGGA 596
Db 181 ArgLysAspGluGlyAspArgGlnGlu-----LysGluGlyArgArgGluGly 196
QY 597 TTTGGGTTCGCGCTCAAGGTGAAGATAATACCTAAAGAGGAACACTGTAAATGCCAGA 656
Db 197 LysGlyArgGlnGluAspGlyArgGluGlyTrp-LysGlu----- 209
QY 657 AGCAGGTGAAGAGCAACACCAA 678
Db 210 ---GlyGluGluGlnGluGln 215
RESULT 13
Q9RV01 PRELIMINARY; PRT; 319 AA.
AC Q9RV01;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 33.1 KDA PROTEIN.
GN DR1229.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RL;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RL.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001971; AAF10810.1; -.
DR TIGR; DR1229; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 319 AA; 33082 MW; 7F0229C5DAAA2B61 CRC64;

Alignment Scores:
Pred. No.: 0.161 Length: 319
Score: 104.00 Matches: 49
Percent Similarity: 39.62% Conservative: 14
Best Local Similarity: 30.82% Mismatches: 57
Query Match: 7.54% Indels: 39
DB: 16 Gaps: 9

US-09-658-824-808 (1-781) x Q9RV01 (1-319)
QY 19 GCGACTCGGTCCTCGAGTCTGGATTCTTCTCCGCTACTGAGACACGCGGGTAGGT 78
Db 159 GlySerSerGlyProSerSerGlyCysSerAlaAsnThrThrArgAlaArgHisAlaPro 178
QY 79 CCACAGCAGATCCAACTCGGAGTTGAAGTGTGAGTGAGAGTGAAGAGGAACACGAGGC 138

```

Best Local Similarity:	28.33%	Mismatches:	81
Query Match:	7.50%	Indels:	28
DB:	11	Gaps:	8

US-09-658-824-808 (1-781) x Q9CSL7 (1-498)

```

QY   162 GACTCAGAGTGAGAAGGCCCTCAAGTCGGTCCCTCTCATCGGTGCGACGCCCATGG 341
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    84 ASplysGluArgGluSerValGluLysArgHisProSerHisProSerProAlaProPro 103

QY   222 ACCTTTCTGTCTCGTCACGCCATAAATAGGGAGGAGGAGGCCGAGGTGGAGGGGC 281
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   104 ValProValSerAlaLeuGlyHisAsn-----ArgSerSerThrAspProThrThr 120

QY   282 TCAGGCGAAGCTGGGTGCTGTTGGGGGTATCCGAGTCCCGAAGCACCTGGAAACC CGGA 341
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   121 ArgGlyHisLeuasn-----ThrGlualaargglulysasplysProLys 135

QY   342 CAGAAGATTCTGGACTCCCAGACGGGACGAGGAGGAGGACGCATGAGCGCACACACA 401
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   136 GlulysGluArgaspHis---SerGlySerArgLysAspLeuThrThrGluGluHisLys 154

QY   402 AACACAGAACCACACACCCAGTCCCAGGAG-----CCCAGTAATGAGAGCCCCAAA 455
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   155 AlaLysGluSerHisLeuProGluuArgaspGlyHisSerHisGluGlyArgAlaAlaGly 174

QY   456 GAAGAACACGAGCTGAAAGTCGGGATCCCT-----ACACTGGCGAGCAG 5000
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   175 GluGluProLysGlnLeuSerArgValProSerProTyrValArgThrProGlyValasp 194

QY   501 ACACAGAAGATCAGGATACAGTCAGATCCCATCGCGGACATGGAAGGTGATCTGCAA 560
      :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   195 SerThrArgProAsnSerThrSerSer-----ArgGluAlaGluProArgLysGly 211

QY   561 GAGCTGCATCAGTCAAAACACCGGGGATFAAATCTGGATTGGGTTCGGCGCTCAAGTGAA 620
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   212 GluProAlaTyrGluAsn-ProLysLysAsnAlaGlu-----VallysVally 227

QY   621 GATAATACCTAAAGAGNACACTGTAATGCCAGAGAGAGGTGMAGAGCAACACAA 678
      | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   227 sGluGluArgLysGluAspHis-----AspLeuProThrGlualaProGln 242

RESULT 15
Q63134 PRELIMINARY; PRT; 246 AA.
AC Q63134
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DI 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CONTIGUOUS REPEAT POLYPEPTIDE PRECURSOR.
GN CRP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus
OX NCBI_TaxID=10116;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE=87165965; PubMed=3558393;
RA Heinrich G., Habener J.F.;
RT "Genes encoding proteins with homologous contiguous repeat sequences
RT are highly expressed in the serous cells of the rat submandibular
RT gland.";
RL J. Biol. Chem. 262:5262-5270(1987).
RW EMBL; M17703; AAA40971.1; -.
KW SIGNAL.
FT SIGNAL.
SO SEQUENCE. 246 AA. 18 POTENTIAL.
   1 18 5BDA8BA0477A92DC CRC64;
```

Alignment Scores:		
Pred. No.:	0.193	Length:
Score:	103.00	Matches:
Percent Similarity:	40.29%	Conservative:
Best Local Similarity:	28.06%	Mismatches:
		50
		246

Query Match:	7.46%	Indels:	33
DB:	11	Gaps:	6
US-09-658-824-808 (1-781) x Q63134 (1-246)			
QY	306	GGGGTATCCGAGTCCCGAAGACACCTGGAAACCCGACAGAGAAGATTCTGGACTCCCGACAG	365
DB	74	GlySerGluGluGluGlnGlnGlnGlnProThrGlnAlaGluAsnGlnGluProPro	93
QY	366	GGGACGAGAGGACGGGATGACGACACACACAAACACAGAACACACAGCCAGTCC	425
DB	94	AlaThr-----SerGlySerGluGluGlnGlnGln	104
QY	426	CAGGAGCCAGT-----AATGGAGAGCCC-----	449
DB	105	GlnGluProThrGlnAlaGluAsnGlnGluProProAlaThrSerGlySerGluGluGlu	124
QY	450	-----CAAAAGAACACACGCA-----GCTGAAAGTCGGGATCTGACCTCCAGTCCAGCAGACA	503
DB	125	GlnGlnGlnGlnGlnProThrGlnAlaGluAsnGlnGluProProAlaThrSerGlySer	144
QY	504	GAAGAGATCAGATACAGTCCAGATCCAGTCCGCGACATGGAAGGTGATCTGCAAGAG	563
DB	145	GluGluGluGlnGlnGlnGlnGlu-----SerThrGlnAlaGluAsnGlnGlu	160
QY	564	CTGCATCAGTCAACACCCGGGATAAATCTGGATTGGTTCCGCGCTCAAGGTCAAGAT	623
DB	161	-----ProSerAspSerAlaGlyGluGlnGluThrGlnProGluGlu	175
QY	624	AATACCTAAGAGGACACTGTAAATCCAGACAGCAGGTGAAGAGCAACACAA	678
DB	176	GlyAsn-ValGluSerProProSerSerProGluAsnSerGlnGluGlnProGln	193
RESULT 16			
ID	Q99MM1	PRELIMINARY;	PRT; 413 AA.
AC	Q99MM1;		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	WSL-1-LIKE PROTEIN.		
GN	TNFRSF12.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=129/SV;		
RX	MEDLINE=21158384; PubMed=11261933;		
RA	Wang E.C.Y., Kilson J., Thern A., Williamson J., Parrow S.N.,		
RA	Owen M.J.;		
RT	"Genomic structure, expression, and chromosome mapping of the mouse		
RT	homologue for the WSL-1 (DR3, Apo3, TRAMP, LARD, TR3, TNFRSF12)		
RT	gene."		
RL	Immunogenetics 53:59-63(2001).		
DR	EMBL; AF329969; AAK11256.1;		
DR	HSSP; P25942; 1CDF.		
DR	MGI; MGI:1934667; Tnfrsf12.		
DR	InterPro; IPR000488; Death.		
DR	InterPro; IPR000561; EGF-like.		
DR	InterPro; IPR001368; TNFR_c6.		
DR	Pfam; PF00020; TNFR_c6; 3.		
DR	SMART; SM00005; DEATH; 1.		
DR	SMART; SM00208; TNFR; 3.		
DR	PROSITE; PS0017; DEATH_DOMAIN; 1.		
DR	PROSITE; PS01186; EGF_2; UNKNOWN_1.		
DR	PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.		
DR	PROSITE; PS00050; TNFR_NGFR_2; 1.		
SQ	SEQUENCE 413 AA; 44453 MW; 69F21B85D0DABAF CRC64;		
Alignment Scores:			
Pred. No.:	0.209	Length:	413
Score:	103.00	Matches:	63
Percent Similarity:	33.98%	Conservative:	24
Best Local Similarity:	24.61%	Mismatches:	70
Query Match:	7.22%	Indels:	99
DB:	11	Gaps:	14
US-09-658-824-808 (1-781) x Q99MM1 (1-413)			
QY	677	TGTGTGCTCTTTCACCTGCTTCTGGCATTTTACAGTGTCTCTTATAGGTATTATCTTC	618
DB	137	CysGlyLysSerSerProPheSer-----CysValProCysGly-AlaThrTh	152
QY	617	ACCTTGAGCGCGGAACCAATCCA--GATTTATCCCGCTGTTT-----GA	573
DB	152	rProValHisGluAlaProThrProArgProCysLeuProGlyPheTyrIleArgGlyAs	172
QY	572	CTGATGACAGCTCTTGCAGATCACCTTCATGTCGCGCACTGGGATCTCAGTGTATCTG	513
DB	172	nAspCysThrSerCys-----ProThrGlyPheSerSerVal-----	184
QY	512	ATCTTCTCTCTGCTGCTGCCCGAGGTGTAGGATCCCGACTTTTCAGCTGCTGTTCTTCT	453
DB	185	-----CysProLysAlaCysThrAlaValCysGlyTrpLysGlnMetPh	199
QY	452	TGGGGCTCTCCATTACTTGGCTCTGCGACTGTGTGTCTGTGTCTGTGTGTGTGTGT	393
DB	199	eTrpValGlnValLeuLeuGlyVal---AlaPheLeuPheGlyAlaIleLeuIleCysAl	218
QY	392	CGCTCATGCCCTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	333
DB	218	a-----TyrCysArgTrpG1	223
QY	332	CAGGTGC-----TTCTGGGACTCGGATACCCCAACAGCACCCCA---GCTTGCCTGA	282
DB	223	nProCysLysAlaValValThrAlaAspThrAlaGlyThrGluProLeuAlaSerProG1	243
QY	281	GCCCTCCACTCCTCGGCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT	222
DB	243	nThAlaHisLeuSerAla-----	249
QY	221	CCATGGGCTGGCACCCGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	162
DB	250	-----SerAspSe	252
QY	161	ACTGACCACACACCCCTCCGGAAGCTGCTGCTGCTCCTCCTCCTCCTCCTCCTCCT	102
DB	252	rAla--HisThrLeuLeuAlaProProSerSerThrGlyLysIleCysThrThrValGln	271
QY	101	CTCCAGTTGGATCT-----	78
DB	272	Leu---ValGlyAsnAsnTrpThrProGlyLeuSerGlnThrGlnGluValValCysGly	290
QY	77	-----CCTACCCGCGCTGCTCAGTAGCGGAGAAAGAA	45
DB	291	GlnAlaSerGlnProTrpAspGlnLeuProAsnArg-----ThrLeuGlyThrProLeu	308
QY	44	TCCAGACTCAGGACCGGAGTCCCGGCTCCAGCTCCCGCC	3
DB	309	AlaSerProLeuSerProAlaProAlaGlySerProAla	322
RESULT 17			
ID	Q9U229	PRELIMINARY;	PRT; 643 AA.
AC	Q9U229;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	Y56A3A.32 PROTEIN.		
GN	Y56A3A.32		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		

RN	[1]	OS	Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
RP	SEQUENCE FROM N.A.	OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
RA	Matthews L.;	OC	Gammapherpesvirinae; Rhadinovirus.
RL	Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.	OX	NCBI_TaxID=37296;
RN	[2]	RN	[1]
RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.
RX	MEDLINE-99069613; PubMed=9851916;	RX	MEDLINE-21165304; PubMed=11264383;
RA	none;	RA	Pilot T., Tramier M., Coppey M., Nicolas J.C., Marechal V.;
RT	"genome sequence of the nematode C.elegans: A platform for	RT	"Close but distinct regions of human herpesvirus 8 latency-associated
RT	investigating biology."	RT	nuclear antigen 1 are responsible for nuclear targeting and binding to
RL	Science 282:2012-2018(1998).	RT	human mitotic chromosomes.";
CC	-1- COFACTOR: FAD (BY SIMILARITY).	RL	J. Virol. 75:3948-3959(2001).
CC	-1- SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES	DR	EMBL; AF305694; AAG27458.1; -.
CC	CLASS-I.	DR	HSP; P02649; INFN.
CC	EMBL; AL132860; CAB60511.1; -.	DR	InterPro; IPR002965; P_rich_extensn.
DR	InterPro; IPR001327; FAD_pyl_redox.	DR	InterPro; IPR002017; Spectrin.
DR	InterPro; IPR001100; Pyl_redox.	DR	InterPro; IPR000533; Tropomyosin.
DR	Pfam; PF00070; Pyl_redox; 1.	DR	PRINTS; PR01217; PRICHEXTENS.
DR	PRINTS; PR00368; FADPNR.	DR	PRINTS; PR00194; TROPOMYOSIN.
DR	PRINTS; PR00411; PNDRTASEI.	FT	NON_TER 1036 1036
KW	FAD; Flavoprotein; Oxidoreductase; Redox-active center.	SQ	SEQUENCE 1036 AA; 119328 MW; 2959EDD2C1C042B8 CRC64;
SQ	SEQUENCE 643 AA; 71257 MW; EAC5A54980A5F75D CRC64;		
Alignment Scores:		Alignment Scores:	
Pred. No.:	0.224	Pred. No.:	0.269
Score:	103.00	Score:	102.50
Percent Similarity:	41.84%	Percent Similarity:	37.78%
Best Local Similarity:	26.24%	Best Local Similarity:	23.08%
Query Match:	7.46%	Query Match:	7.43%
Indels:	26	Indels:	46
Gaps:	8	Gaps:	4
DB:		DB:	
US-09-658-824-808 (1-781) x Q9U229 (1-643)		US-09-658-824-808 (1-781) x Q9DUM3 (1-1036)	
QY	207 GTGCCAGCCCA-----TGACCTTCTGCTCGTCACGGCCATACTAGGGA 254	QY	261 AGGGCCGAGAGTGGAGGGCTCAGCGAGCTGGGGTGTGGGTATCCGAGTCC 320
DB	45 IleProLysProGlySerLeuAspTrpThrPheSerArg---SerHisThrLysSer 63	DB	432 LysLysGluAspGluGluAspGlyGlyAspGlyAsnLysThrLeuSerIleGlnSerSer 451
QY	255 GGAAGGAGGCGCAGGAGTGGAGGGCTCAGCGGAAGCTGGGTGCTGTTGGGG----- 308	QY	321 CAGAACCACTCGAACCCGACACAGAGATTCTGGACTCCCGACAGCGGACGAGGAGGG 380
DB	64 AlaHisGluPheGluProTyrLys-----ProGluIleGlyAlaPheIleGlyAlaVal 81	DB	452 GlnGlnGlnGlnGluProGlnGlnGln----- 460
QY	309 -----GTATCCGAGTCCCAAGAGCAGCTGGAAACCCGACAGAGATTCTGGACTCCCCA 362	QY	381 ACGGCATGAGCGACACACACAAACACAGACACACAGC-----CAGTCCCAGGAG 431
DB	82 AlaPheIleGlyLeuThrLeuIleAlaValIleLysThrAspValPheLysLysGlu 101	DB	461 -----GluProGlnGlnGlnGluProGlnGlnGlnGlu 471
QY	363 GACGGACAGGAGGAGCGGATGAGCGACACACACA-----AACACAGAACACAC 416	QY	432 CCAGTAATGGAGAGCCCCAAAAGAACACAGAGCTGAAAGTCGGATCCTACACCT 491
DB	102 AspSerHisGlyHisGlyHisGlyHisAlaLysHisSerLysLysHisGluGluLysHis 121	DB	472 ProGlnGlnGlnGluProGlnGlnGlnGluProLeuGlnGlnGlnGlnGlnGluPro 491
QY	417 AGCCAGTCC---CAGGAGCCCACTAATGGAGAGCCCCAAAAGAACACACAGCA----- 467	QY	492 GGCACGACAGACAGAAAGATCAGGATACAGTCCAGATCCCATGCCGACATGGAAGGT 551
DB	122 GluGlnLysHisGluGluLysGluHisAlaGluProGluLysLysGluGluAlaLysPro 141	DB	492 GlnGlnGlnGluProGlnGlnGlnGluProGlnGln-----GlnGlu 505
QY	468 -----CCTGAACTCGGATCCTACCTGGGCG----- 497	QY	552 GATCTCAAGAGCTGCATCAGTCAAAACACCGGGGATAAATCTGGATTGGGTTCCGCGGT 611
DB	142 GluLysProAlaGluProLysGluProGluProGluProAlaGlnLysGlnAlaGluGlnProGlu 161	DB	506 ProLeuGlnGluProGlnGln-GlnGlu----- 514
QY	498 CAGACAGAAGATCAGGATCAGATCCAGTCCAGTCCGCGACATGGAAGGTGATCTG 557	QY	612 CAAGGTGAAGATAATACCTAAAGAGAACACTCTAAATGCCAGAGAGGTAAGAGCA 671
DB	162 GlnAlaGluGluLysGlnGluThrLysAspAlaGluProLysGluGlnValAspAspArg 181	DB	515 -----ProGlnGlnGlnGlnGluProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 529
QY	558 CAA 560	QY	672 ACCACAA 678
DB	182 Gln 182	DB	529 uProGln 531
RESULT 18		RESULT 19	
Q9DUM3	PRELIMINARY; PRT; 1036 AA.	Q98148	PRELIMINARY; PRT; 1162 AA.
ID	Q9DUM3	ID	Q98148
AC	Q9DUM3; 2001 (TrEMBLrel. 16, Created)	AC	Q98148; 01-FEB-1997 (TrEMBLrel. 02, Created)
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)	DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	LATENT NUCLEAR ANTIGEN (FRAGMENT).	DE	ORF73 HOM <sup>1</sup> 95.
		OS	Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).



Db	1346	ProLeuAlaProLeu	:	:	:	:	ProValLeu-AlaProSerProGl	1358
Qy	418	CTGTGTG	---GTTCTGTGTTGTGTCGTGCTCATGCGTCCCTCTCCCTGGTCCCGCTCT	362				
Db	1358	YAlaAlaProValLeuAlaSerSerGlnThrProValProValMet	:	:	:	:		1373
Qy	361	GGGAGCTCCAGAACTCTCTGTGCGGGTTCACAGGTCTCTCGGACTCGGATACCCCAAC	302					
Db	1374	AlaProSerSerThrProGlyThrSerLeuAlaSerProValPro	:	:	:	:		1390
Qy	301	AGCACCAGCTTCGCTGAGCCCTCCACT	263					
Db	1391	AlaProThrProValLeuAlaProSerSerThrGlnThrMetLeuProAlaProValPr	1410					
Qy	262	CTCTTCTCCCTAGTATGCGCGTCACGACACAGAAGTCCATGGCGTGGCACCGCA	203					
Db	1410	oSerProLeuProSerProAlaSerThrGlnThr	1427					
Qy	202	TGAGAGGAGCAGGACTTCGAGGGCTTCTCACTCTGAGTCACTG	152					
Db	1427	laLeuAlaProThr-LeuGlyGlySerProSerGlnThrLeuSerLeuGlyThrGly	1446					
Qy	151	CAACCTCCGGAGCGCTGCTCTCTCACTCTCACTC	111					
Db	1447	AsnProGlnGly--ProPheProThrGlnThrLeuSerLeuThrProAlaSerSerLeuV	1466					
Qy	110	-----ACACTTCAACTCCAGTTGGATCTGCTGTGGACCTACCGCC	68					
Db	1466	alProThrProAlaGlnThrLeuSerLeuAlaProGlyProProLeuGlyProThrGlnT	1486					
Qy	67	GTCTCTCAGTAGGAGGAG	20					
Db	1486	hrLeuSerLeuAlaProAlaProProLeuAlaProAlaSerProValGlyPro--Alap	1505					
Qy	19	CGGCTCACAGC	9					
Db	1505	roAlaHisThr	1508					
RESULT	21							
Q33572		PRELIMINARY;	PRT;	350	AA.			
ID	Q33572							
AC	Q33572;							
DT	01-NOV-1996 (TrEMBLrel. 01, Created)							
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)							
DE	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)							
DE	KINETOPLAST APOCYTOCHROME B.							
GN	COB OR CYTB.							
OS	Trypanosoma brucei.							
OG	Mitochondrion.							
OC	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.							
OX	NCBI_TaxId=5691;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RA	MEDLINE=85086128; PubMed=6513991;							
RA	Johnson B.J.B., Hill G.C., Donelson J.E.;							
RT	"The maxicircle of Trypanosoma brucei kinetoplast DNA encodes							
RL	apocytochrome b.,"							
RL	Mol. Biochem. Parasitol. 13:135-146(1984).							
CC	-1- FUNCTION: COMPONENT OF THE UBIOUINOL-CYTOCHROME C REDUCTASE							
CC	COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A							
CC	RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL							
CC	COUPLED TO ATP SYNTHESIS (BY SIMILARITY).							
CC	-1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY							
CC	BOUND TO THE PROTEIN (BY SIMILARITY).							
CC	-1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B6-F ARE: CYTOCHROME B6, 17							
CC	KDA POLYPEPTIDE (PETD), CYTOCHROME F AND THE RIESKE PROTEIN (BY							
CC	SIMILARITY).							
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.							
DR	EMBL; M17998; AAA32115.1; -							
DR	InterPro; IPR000179; Cyt_b.b6.							
DR	Pfam; PF00033; cytochrome_b_c.1.							
DR	Pfam; PF00033; cytochrome_b_n.1.							

[illegible]

```
Db 436 SerGluAlaGluAspProAlaAlaGluAlaLeuArgArgGluSerAspArgArgGlu 455
QY 531 -----CCAGTGGCGGACATGGAAGTGATCTGCAAGAGCTGCATCAGTCAAAACACC 581
Db 456 AlaAlaAspArgLeuArgGluAlaGluGluThrAlaGlnGluAlaAlaArgAla 473
QY 582 GGGGATAAATCTGGATTCTGGTTCGGCGCTCAAGGTGAAGATAAATACCTAAAGAGGAACA 641
Db 474 -----ArgGlnAlaGluGluThrAla-ArgGluGluAl 484
QY 642 CTGTAATAATCCAGACAGAGGTCAA 666
Db 484 aAlaArgAlaArgGlnAlaGlyGlu 492
RESULT 24
Q27033
ID Q27033 PRELIMINARY; PRT; 480 AA.
AC Q27033;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE MEMBRANE PROTEIN.
OS Theileria parva.
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
OC Theileria.
OX NCBI_TaxID=5875;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MUGUGA;
RC MEDLINE=94088665; PubMed=8264721;
RA Baylis H.A., Allsopp B.A., Hall R., Carrington M.;
RT "Characterisation of a glutamine- and proline-rich protein (QP
RL Mol. Biochem. Parasitol. 61:171-178(1993).
DR EMBL; L06323; AAA18800.1; -.
SQ SEQUENCE 480 AA; 52388 MW; 619B46AE128EA2E3 CRC64;
```

```
Alignment Scores:
Pred. No.: 0.417 Length: 480
Score: 100.00 Matches: 40
Percent Similarity: 43.66% Conservative: 22
Best Local Similarity: 28.17% Mismatches: 42
Query Match: 5 Indels: 38
DB: 5 Gaps: 8
```

US-09-658-824-808 (1-781) x Q27033 (1-480)

```
QY 341 ACAGAAAGATTCTGGACTCCCGACGGGACAGGAGGCGGATGAGCGACACAC 400
Db 68 ThrSerGlnGlnGlyLysProGlnProAspGlnProGlnAsp- --GlnProAspGlnHls 86
QY 401 AAACAC-----AGAACCCACAGCCAGTC 424
Db 87 GlnGlnProThrGlnGlyAspThrSerGlyGlnGlnGlyProAspThrProGlnProIle 106
QY 425 CCAGGAGCCAGTAATGGAGAGCCCCCAAAAGAA-----GAACC 463
Db 107 -GlnGluProSer---GlyProValGlnProAspGlnThrGlyGlnGlyProValGluPr 125
QY 464 AGCAGCTGAAATCGGGATTCCT-----ACACTGGGCGACGACAGAGAAGA 511
Db 125 oValAspGlnGlnGlnProThrGlnGlyAspThrSerGlyGlnGlnGlyGlnGlnPr 145
QY 512 TCAGGATACAGTGAGATCCCGACATGGCAATGGATGATCTGCAAGAGCTGCATCA 571
Db 145 oGlnAspGln-----ProValGlnGlnAspGlyGlnAspSerGlnGlyThrPr 162
QY 572 GTCAACACCGGGGATAAATCTGGATTGGGTTCGGGGTCAAGGTGAAGATAATACCT- 630
Db 162 oGluGlnThrProAspGlnSerGly-----GlnGlnProGlyProAspThr-PROA 179
QY 631 -----AAAGAGGAACACTGTAAATCCCAAGACAGGAGGTGAAGAGCAACACC 676
```

```
Db 179 spGlnProValTyrGlnGlnGlnProValGlnGlnProSerGlyGlnGlnGlnProG 199
QY 677 AA 678
Db 199 ln 199
RESULT 25
Q94E85
ID Q94E85 PRELIMINARY; PRT; 181 AA.
AC Q94E85;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE B1045D11.17 PROTEIN.
GN B1045D11.17.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
clone:B1045D11.17";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003199; BAB61149.1; -.
SQ SEQUENCE 181 AA; 19391 MW; 4B7CDC5D082442AF CRC64;
```

```
Alignment Scores:
Pred. No.: 0.401 Length: 181
Score: 99.50 Matches: 49
Percent Similarity: 43.68% Conservative: 27
Best Local Similarity: 28.16% Mismatches: 58
Query Match: 7.21% Indels: 40
DB: 10 Gaps: 10
```

US-09-658-824-808 (1-781) x Q94E85 (1-181)

```
QY 164 CTCAGAGTGAGAAAGCCCTCGAAGTCGTCCCTCTCA-----TGGCGTGCCACG 214
Db 7 lleArgTPrArgArgProAlaAlaLeuSerProLeuSerProAlaGlySerGlySerPro 26
QY 215 CCATGGACCTTCTTCTCTCGTCAGGCCATAACTAGGAGGAGGAGGCGGAGGAGTG 274
Db 27 ArgLeuLeuPheLeuProSerProAspPro-----AlaGluGlyArgGlyVal 42
QY 275 GAGGGCTCAGCGGAAGCTGGGTGCTGTT-----GGGGGTATCGGAGTCCCA 322
Db 43 AspGly-----GlyAlaValGluGlyGluProGlyGlyGlnArgGluHls 57
QY 323 GAAGCACCTTGA-----ACCCGACAGAAAT-----TCTGGACTCCCCAGA 364
Db 58 GlySerGlyGlyArgProLeuSerProThrArgAspArgAlaSerAlaProLeuProSer 77
QY 365 CGGACACAGGAGAGCGGATGAGCGCACACACAAACACACACACACAGCAATC 424
Db 78 ProProLeuProArgSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 95
QY 425 CCAGGAGCCACAG---TAATGAGAGCCCCCAAAAGAAAGAACCA-----464
Db 96 ArgGlyAlaArgSerArgTrpArgAlaValArgArg- ProProProThrMetThrArgTh 115
QY 465 -----GCAGCTGAAGCTCGGATCTTACCTCGGCGACGACAGACAGAAAGATCAGGA 517
Db 115 rGlyValAlaSerSerAlaLysAspGluAspAspGluGluAspAspLysGluGlu 135
QY 518 TACAGCTCAGATCCAGTCCGCGCAGATGGAAGGTGATCTGCAAGAGCTGCATCAGTCAAA 577
Db 135 uAlaAlaAlaProProAsnSerArgArg-----TyrGluSerAl 149
```



```

Qy 578 CACCCGGGGATAAATCTGGATTTCGGTTCGGCGCTCAAGGT 617
Db 149 alleSerAspLeuAlaGlyHisGlypHeGluGluGly 162
      |||  |||  |||  |||  |||  |||  |||  |||
      |||  |||  |||  |||  |||  |||  |||  |||
RESULT 26
ID Q15627 PRELIMINARY; PRT; 325 AA.
AC Q15627;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TR3BETA.
OS Homo sapiens (Human).
CS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N. A.
RP TISSUE=SKLETAL MUSCLE;
RA Okura N., Ito M., Tsukada T., Sasaki K., Yamaguchi K., Miki K.;
RT "Identification of an isoform of human TR3 (NGRI-B, Nur77).";
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR ( BY SIMILARITY ).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL: D85245; BAA12746.1; -.
DR HSSP: P20393; 1A6V.
DR InterPro: IPR001628; zf-C4.
DR Pfam: PF00105; zf-C4; 1.
DR SMART: SM00399; ZnF_C4; 1.
DR DNA-binding: Nuclear protein; Receptor; Transcription regulation;
DR Zinc-finger.
SQ SEQUENCE 325 AA; 33754 MW; BCF2490A32A09175 CRC64;

```

Alignment Scores:		
Pred. No.:	0.439	Length: 325
Score:	99.50	Matches: 56
Percent Similarity:	37.11%	Conservative: 16
Best Local Similarity:	28.87%	Mismatches: 72
Query Match:	6.98%	Indels: 50
DB:	4	Gaps: 7
US-09-658-824-808 (1-781) x Q15637 (1-325)		

568	QY	TGCAGCTCTTGCAGATCACACTTCCATGTCTCGCGCACTGGGATCTCAGCTGATCTATCCTGATCT	509
78	Db	CysSerAlaSerSerAlaSerSerThrSerSerSerAlaThrSerProAla	97
508	QY	TCTTCTCTC-----TGCTGCCCAAGT-----	488
98	Db	SerAlaSerPheLysPheGluAspPheGlnValTyrGlyCysTyrProGlyProLeuSer	117
487	QY	-----CTAGGATCCCGACTTTTCAGCTGCTGGTCTCTCTTTTGGGGCTCCCATTA---	437
118	Db	GlyProValAspGluAlaLeuSerSerGlySerAspTyrTyrGlySerProCysSer	137
436	QY	-----CTGGGCTCTCGGACCTGGCTG	416
138	Db	AlaProSerProSerThrProSerPheGlnProProGlnLeuSerProThrPaspGlySer	157
415	QY	TGTGGTTCTGTGTTTGTCGTGCGCTCANGCGGTCCTCTCTGCTCCGCTCTGGGGAG	356
158	Db	PheGly-----HisPheSerPro-SerGlnThrTyr-GluGlyLe	170
355	QY	TCCAGATCTTCTGTCGGGTTCCAGGTGCTCTCTGGACTCGGATACCCCAACAGCAC	296
170	Db	uArgAlaTrpThrGluGlnLeuProLysAlaSerGly-----ProPro-GlnProP	187
295	QY	CCAGCTTCGCGCTGAGCGCCCTCCACTCTCGGCGCTCCTTCTCCCTAGTATTATGGCGGTGA	236
187	Db	roAlaPhePheSerProProThrGlyProSerProSerLeuAla	203
235	QY	CGAGACAAGAAGTCCATGGCGGTGGCCACCGCATGAGAGGACGACGACTTCGAGGGCCT	176

Db	204	-----GlnSerProLeuLysLeuPheProSerGlnAlaThr-HisGlnLeuGlyGlu	229
Qy	175	TCTCACTCTGAGTCACGTACCACACACCCCTC;GGAAAGCGTGGTGTTCTCTTCACTCT	116
Dd	221		111
Dd	221	GlyGluSerIyrSerMetProThrAlaPheProGly-----LeuAlaProThrSerPro	238
Qy	115	CACTCACACITCAACTCCCAGTTGGATCTGCCTGTG	80
Dd	239	HISLeuGluGlySerGlyIleLeuAspThrProval	250
		:	
<b>RESULT 27</b>			
QyUEU5	ID	QyUEU5 PRELIMINARY;	PRT; 116 AA.
AC	QyUEU5		
DT	01-MAY-2000	(TREMBLrel. 13, Created)	
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	GAGE"-8.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=93323388; PubMed=10397259;		
RE	De Backer O., Arden K.C., Boretti M., Vantomme V., De Smet C.,		
RA	Czekay S., Viars C.S., De Plaen E., Brasseur F., Chomez P.,		
RA	Van den Eynde B., Boon T., van der Bruggen P.;		
RT	"Characterization of the GAGE genes that are expressed in various		
RL	human cancers and in normal testis.";		
RL	Cancer Res. 59:3157-3165(1999).		
DR	EMBL; AF055473; AAC33676.1; -.		
SQ	SEQUENCE 116 AA; 12764 MW; DD30552933966F19A CRC64;		

Alignment Scores:		
Pred. No.:	0.419	Length: 116
Score:	99.00	Matches: 38
Percent Similarity:	39.71%	Conservative: 16
Best Local Similarity:	27.94%	Mismatches: 59
Query Match:	7.17%	Indels: 23
DB:	4	Gaps: 4

```

US-09-558-824-808 (1-781) x Q9UEU5 (1-116)
Qy 273 TGGAGGGGCTCAGCGCAAGCTGGGTGCTGTGGGGGTATCCGAGTCCCAAGAAGCACCTG 332
|||||
Db 3 TrpArgGlyArgSerThrTyrArg-----ProArgProArgArgTyrVal 17
|||||
Qy 333 GAACCCGCACAGAGATTCTGGACTCCCCACAGCGGACAGGAGAGAGCGGACGGCATGAGCG 392
|||||
Db 18 GluProProGluMetIle-----GlyProMetArg----- 27
|||||
Qy 393 ACACACAAACACAGAAACACACAGCCAGCTCCCGAGGACCCAGTAAATGGAGAGCCCAA 452
|||||
Db 28 -----ProGluInPheSerAspGluValGluProAlaThrProGlu 41
|||||
Qy 453 AAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACTGGGCGAGCAGACAGAAGAAGAT 512
::: ||||||| ::: |||||||
Db 42 GluGlyGluProAlaThrGlnArgGlnAspProAlaAlaGlnGluGlyGluAspGlu 61
|||||
Qy 513 CAGGATACAGCTGAGATCCCGACTGCGGACATGGAAGTGTATCGCAAGAGCTGCATCAG 572
|||||
Db 62 GlyAlaSerAlaGlyGlnGlyProLysPro---GluAlaAspSerGlnGluGlnGlyHis 80
|||||
Qy 573 TCAACACACCGGGGATAAATCTGGATTTCGGGTTCGGCGCTCAAGGTGAAGATAATACCTAA 632
|||||
Db 81 ProGlnThrGlyCysGluCysGluAspGlyProAspGlyGlnGluMetaspPro-ProAs 100
|||||
Qy 633 ACAGGAACACTGTAAATGCCAGAACGAGGTGAAGACCAACCACAA 678
|||||
Db 100 nProCUGluValLysThrProGluGlyGluLysGlnSerGln 115
|||||
RESULT 28

```

```
Q99JV6          PRELIMINARY;          PRT;      630 AA.
ID  Q99JV6
AC  Q99JV6;
DT  01-JUN-2001 (TrEMBLrel. 17, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  HYPOTHETICAL 67.2 KDA PROTEIN (FRAGMENT).
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Strausberg R.;
RL  Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; BC005641; AA005641.1; -.
KW  Hypothetical protein.
FT  NON_TER
SQ  SEQUENCE 630 AA; 67197 MW; 58DD514CADBAE492 CRC64;

Alignment Scores:
Pred. No.:      0.543      Length:      630
Score:          99.00      Matches:     44
Percent Similarity: 37.14%      Conservative: 21
Best Local Similarity: 25.14%      Mismatches: 70
Query Match:      7.17%      Indels:    40
DB:              11      Gaps:      9

US-09-658-824-808 (1-781) x Q99JV6 (1-630)
QY 138 CTTCCGGAGGTTGTGGTCAGTCAC-----TCAGAGTGAGAGGCCCTCGAAGTCGT 191
Db 49 ILeProHisAspLysValValAspGluGlnAspGluGlyProGluGluLys 68
QY 192 CGTCCCTCATCGGGTGCACGCCACGCCCTTCTGTCTCGTCACGCCATAACTAG 251
Db 69 LysProProArgLeuProAspGlu-----
QY 252 GGAGGAGGAGGCGGAGGAGTGGAGGGCTCAGCGCAAGCTGGGTGCTGTGGGGGTA 311
Db 78 --GlyAspProAla-----GlyArgGlyGlnGlyAlaProProLeuProGlu 92
QY 312 TCCAGAGTCCCAAGACACCTGGAAACCCGACAGAGATCTCGACTCCCGAGCGGACC 371
Db 93 SerGluLysGluLysGln--GluProGluArg-----GlyGlyGluGlyLys 107
QY 372 AGGAGAGGAGCGGATGAGCGACA-----CACACAACACAGAACCCACAGC 419
Db 108 ArgProGlyGlnValLeuAlaValGlyGluThrgluHisProGlnLysValProGluAla 127
QY 420 CAGTCCCAGGAGCCAGTAATGAGAGCCCAAAAGAACACAGCAGCTGAAGTCGG 479
Db 128 AsnGlyGlnProProVal-----GlnProArgLysGluAspSerArgProGlyAsnArg 145
QY 480 GATCCTACACTGGCGGACAGACAGAGAAGAT-----CAGGAT 518
Db 146 AspProGlnProAlaAlaGlnAlaArgAspSerValGluLeuLysAlaLeuAlaAsp 165
QY 519 ACAGCTGAGATCCCGAGCCGACATGGAAGGTGATCTG-----CAAGAGCTG 566
Db 166 AspGlyArgGluProAlaGlnGlnLysAlaGlyAlaLeuTrpLysProValGluSerAla 185
QY 567 CATCAGTCAAAACCGGGGATAAATCTGGATTTGGTTCGGCGGT 611
Db 186 AlaGluSerAspAlaGlyGlyLysAlaGlyLeuProValGlnArg 200

RESULT 29
Q76871          PRELIMINARY;          PRT;      773 AA.
ID  Q76871
AC  Q76871;
DT  01-NOV-1998 (TrEMBLrel. 08, Created)
DT  01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT  01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

EG:100G7.2 PROTEIN.
CG3591 OR EG:100G7.2.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Murphy L., Harris D., Barrell B.;
RL  *Sequencing the distal x chromosome of Drosophila melanogaster.*;
RL  Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RA  Benos P.;
RL  Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AL031366; CAA20518.1; -.
DR  FlyBase; FBgn0014097; CG3591.
SQ  SEQUENCE 773 AA; 86307 MW; 1D073855D5E7A983 CRC64;

Alignment Scores:
Pred. No.:      0.561      Length:      773
Score:          99.00      Matches:     34
Percent Similarity: 42.74%      Conservative: 19
Best Local Similarity: 27.42%      Mismatches: 54
Query Match:      7.17%      Indels:    17
DB:              5      Gaps:      4

US-09-658-824-808 (1-781) x Q76871 (1-773)
QY 221 GACCTTCTGTCTGTACGGCCATACTAGGAGGAGGAGGCGCGAGGTGAGGGG 280
Db 658 AspLeuLeuThrglnAsnArgArgPheAlaLeuArgLysThrglyArgSerLeuGluIle 677
QY 281 CTCAGCGAAGCTGGGTGCTGTGGGGTATCCGAGTCCCAAGCACCTGGAAACCCG 340
Db 678 ProGluGlnValAlaProLysGlyAspLeuGlnLysGluGluLysProLysGluGlu 697
QY 341 ACAGAAGATCTCGATCCCGAGCGACCGAGGAGGAGCGCATGAGCGACACAC 400
Db 698 GluGlnLysGluLysLeuProLysGluGluValGln-----LeuGluGluLys 714
QY 401 AAACACAGAACCCACACAGC-----CAGTCCCAGGAGCCCAAGTAAATGGAGCCCA 451
Db 715 LysGlu-GluProGlnLysGluGluLysGluGluProGlnLysGluGluProGln 734
QY 452 AAAAGAAGAACACAGCAGCTGAA-----AGTCGGGATCTCTACACCTGGCAGCAGACA 505
Db 734 nLysGluGluProArgLysGluGluProGlnLysGluGluProGlnLysGluGluProGln 754
QY 506 AGAAGATCAGGATACAGCTCAGATCCCGAGTCCCGAGTGGAGGTGATCTGCAAGAGCT 565
Db 754 nLysGluGluProLysValGluThrPro-----GlnProLe 766
QY 566 GCATCAGTCA 575
Db 766 uGluGlnSer 769

RESULT 30
Q9W4V0          PRELIMINARY;          PRT;      773 AA.
ID  Q9W4V0
AC  Q9W4V0;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE  ANON-3CB PROTEIN.
GN  CG3591 OR ANON-3CB OR EG:100G7.2.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.M., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaisi M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nixon K., Nusskern D.R., Pacleb J.M., Reese M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrlkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT *The genome sequence of Drosophila melanogaster.*;
RL Science 287:2185-2195(2000).
DR EMBL; AE003425; AAF45836.1; -
DR Flybase; FBgn0014097; CG3591.
SQ SEQUENCE 773 AA; 86247 MW; DFE6A6A9166469BC CRC64;

Alignment Scores:
Pred. No.: 0.561 Length: 773
Score: 99.00 Matches: 34
Percent Similarity: 42.74% Conservative: 19
Best Local Similarity: 27.42% Mismatches: 54
Query Match: 7.17% Indels: 17
DB: 5 Gaps: 4

US-09-658-824-808 (1-781) x Q9W4V0 (1-773)
QY 221 GACCTCTGTGTCGTCACGCCATACTAGGAGGAGGAGGCGCGAGGAGTGGAGGG 280
|||||
Db 658 AspLeuLeuThrGlnAsnArgPheAlaLeuArgLysThrGlyArgSerLeuGluIle 677
|||
QY 281 CTCAGGCGAAGCTGGGGTGTGTTGGGTATCCGAGTCCAGATCCAGACCTGGAAACCCG 340
|||
Db 678 ProGluGluInValAlaProLysGlyAspLeuGluInLysGluGluProLysGluGlu 697
|||
QY 341 ACAGAAGATCTGGACTCCCGACGGGACCGAGGAGGAGGCGCATGAGCGACACACAC 400
|||
Db 698 GluGlnLysGluLysLeuProLysGluGluValGln-----LeuGluGluIleLys 714
|||
QY 401 AAACACAGAACCCACACAGC-----CAGTCCCGAGGCGCCAGTATGAGAGGCCCA 451
|||
Db 715 LysGlu-GluProGlnLysGluGluLeuLysGluGluProGlnLysGluGluProGln 734
|||
QY 452 AAAAGAGAACACAGCAGCTGAA-----AGTCGGGATCCTACACCTGGCGAGCAGACAG 505
|||||
```

[illegible]

Alignment Scores:  
Pred. No.: 0.568 Length: 405  
Score: 98.50 Matches: 46  
Percent Similarity: 32.08% Conservative: 22  
Best Local Similarity: 21.70% Mismatches: 73  
Query Match: 7.14% Indels: 72  
DB: 10 Gaps: 6

US-09-658-824-808 (1-781) x Q95Y0 (1-405)

QY 61 GAGACACGGCGGTAGGTCCACAGGAGATCCAACTGGAGTTGAAGTGTGAGTGAGAGT 120  
Db 50 GlnLeuArgAspValSerProGluCysGlnProValGlyGlyGlyProValAlaArgGln 69  
QY 121 GAAGAGAAACAGCAGGCTCCGGAGGGTGTGTGTCAGTCACTCAGAGTGAGAGGCC 180  
Db 70 TyrGluGln----- 72  
QY 181 CTCGAAGTCGTCTCCCTCTCATCGGTGCCAGGCCCATGGACCTTCTTGTCTCGTCACG 240  
Db 73 ---GlnValValPro----- 77  
QY 241 GCATAACTAGGAGGAGAGGAGGCGGAGAGTGGAGG----- 279  
Db 78 -----ProLysGlyGlySerPheTyrProGlyGluThr 89  
QY 280 -----GCTCAGCGAAGCTGGGTCTGTGGGGTATCCGAGT----- 318  
Db 90 ProProGlnGlnGlnGlnGlnSerIleLeuTrpGlyLeuProAlaLeuArgArgTyr 109  
QY 319 CCAGAACGACCTGGAAACCCGACAGAGATCTGGACTCCCGACAGCGGA----- 369  
Db 110 TyrLeuSerValThrSerProGlnArgValSerTyrTyrProGlyGlnAlaSerSerGln 129  
QY 370 -----CCAGAGAGGACGCGCATGACGACACAC-ACAAACACA 407  
Db 130 ArgProGlyGlnGlnGlnGlnProGlyGlnGlnGlnGlnGlnTyrTyrLeuThrSerPro 149  
QY 408 GAACACACAGCCAGTCCAGGAGGCCAGTAAATGAG----- 446  
Db 150 GlnGlnSerGlyGlnTrpGlnGlnProGlyGlnGlnGlnSerGlyTyrTyrProThrSer 169  
QY 447 CCCCCAAAAGAAACACAGCAGCTGAAAGTGGGATCTTACACCTGGCGAGCAGACAGAA 506  
Db 170 ProGlnGlnSerGlyGlnGlnGlnProGlyTyrTyrProThrSerProTrpGlnProGlu 189  
QY 507 GAAGATCAGGATACAGTCCAGTCCAGTGGCGGACATGGAAGGTGATCTGCAAGAGCTG 566  
Db 190 GlnLeuGlnGlnProThrGlnGlnGlnGlnGlnGlnGlnProGlyGlnGlnGlnLeu 209  
QY 567 CATCAGTCAACACCGGGGATAAATCTGGATTGGG 602  
Db 210 ArgGlnGlnGlnGlnGlnSerGlyGlnGly 221

RESULT 34  
Q95KP7  
ID Q95KP7 PRELIMINARY; PRT; 847 AA.  
AC Q95KP7;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE TREACLE (FRAGMENT).  
GN TCOF1.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Haworth K.E.;  
RT "Canine TCOF1: Cloning, chromosome assignment and genetic analysis in  
RT dogs with different face and head types.\*";

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ302713; CAC88122.1; JOINED.  
DR EMBL; AJ302714; CAC88122.1; JOINED.  
DR EMBL; AJ302715; CAC88122.1; JOINED.  
DR EMBL; AJ302716; CAC88122.1; JOINED.  
DR EMBL; AJ302717; CAC88122.1; JOINED.  
DR EMBL; AJ302718; CAC88122.1; JOINED.  
DR EMBL; AJ302719; CAC88122.1; JOINED.  
FT NON\_TER 847  
SQ SEQUENCE 847 AA; 85646 MW; C5A763PFD1CA2235 CRC64;

Alignment Scores:  
Pred. No.: 0.636 Length: 847  
Score: 98.50 Matches: 55  
Percent Similarity: 40.10% Conservative: 22  
Best Local Similarity: 28.65% Mismatches: 74  
Query Match: 7.14% Indels: 41  
DB: 9 Gaps: 9

US-09-658-824-808 (1-781) x Q95KP7 (1-847)

QY 19 GCGACTCGGTCCTCGAGTCTGGATCTTCTCCGTACTGAGACACGCGGGTAGGT 78  
Db 245 GlyLysAlaGlyPro-----SerAlaAlaGlnAlaLysThrGluLys 258  
QY 79 CCACAGCAGATCCAACTGGAGTGAAGTGTGAGTGAGTGAAGAGAACACGAGGC 138  
Db 259 ProLysGluAspSerAspSerSerGluGluAspSerGluGluProAla 278  
QY 139 TTCGAGGGTGTGTGGTCAGTCACTCAGAGTGAGAGGCCCTCGAAGTCGTCCTCCT 198  
Db 279 AlaLysThrProLeu---GlnValLysProSerGlyLysThrProGlnValLysAlaAla 297  
QY 199 CTCATCGGTGCCACGCCCATGGACCTTCTGTCTCGTCACGGCCATPACTAGGAGGAA 258  
Db 298 SerAlaSerAla---LysGluSerProArgLysGlyValProProValProGlyLys 316  
QY 259 GGAGGCCCGAGTGGAGGGCTCAGCGAAGCTGGGTGCTGTTGGGGGTATCCGAGT 318  
Db 317 ValGlyProAlaAlaGly-----GlnAlaLysLys-GlyAla-----G1 329  
QY 319 CCAGAGGACCTGGAAACCCGACAGAGATCTTGGACTCCCGAGC----- 365  
Db 329 yGluGluAspProAspSerSerThrGluGluSer-AspSerGluGluGluAlaProThrA 349  
QY 366 -----GGGACAGGAGGAGCGCATGAGCGACACACACAAACACACACACACA 417  
Db 349 laValProProThrArg-----SerProValGlnAlaLysProSerG 363  
QY 418 GCCAGTCCCGAGG-----CCCGATATGGAGAGAGCCCAAGAG 459  
Db 363 yGlnAsnSerGlnValArgThrAlaSerGlyProValLysGlyProGlnLysAlaG 383  
QY 460 AACACAGCAGTGAAGTCCGGATCCCTACACCTGGCGAGCAGACAGAGAGATCAGGATA 519  
Db 383 lyProAlaAla-----ThrProValGlyLysGlnGluGluAspSerGluS 398  
QY 520 CAGCTGAGATCCCGAGTCCCGACATGGAAGT 551  
Db 398 erSerSerGluGluGluSerAspSerGluGly 408

RESULT 35  
Q949B8  
ID Q949B8 PRELIMINARY; PRT; 1012 AA.  
AC Q949B8;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 109.0 KDA PROTEIN.  
GN C101SERIPDK.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```
OC Ehrhartoidae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21329048; PubMed=11435398;
RA Mayer K., Murphy G., Tarchini R., Mambutt R., Volckaert G., Pohl T.,
RA Duesterhoeft A., Stiekema W., Entian K.D., Terryn N., Lemcke K.,
RA Haase D., Hall C.R., van Dodeweerd A.M., Tingey S.V., Mewes H.W.,
RA Bevan M., Bancroft I.;
RT *Conservation of microstructure between a sequenced region of the
RT genome of rice and multiple segments of the genome of Arabidopsis
RT thaliana.*;
RL Genome Res. 11:1167-1174(2001).
DR EMBL; AJ307662; CAC39074.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1012 AA; 108951 MW; 4BCE3C788BDC44D CRC64;

Alignment Scores:
Pred. No.: 0.653 Length: 1012
Score: 98.50 Matches: 54
Percent Similarity: 34.33% Conservative: 15
Best Local Similarity: 26.87% Mismatches: 84
Query Match: 7.14% Indels: 48
DB: 10 Gaps: 8

US-09-658-824-808 (1-781) x Q949B8 (1-1012)
QY 123 AGAGAACACAGCGCTCCGGAGGTGTGTGCTCAGTCAGTCAGAGTCAGAGAGGCCCT 182
    ||||| ||||| ||||| ||||| |||||
Db 321 ArgGlyGluGlnSerIleProArgGlyGlyArgAlaSerGlySerArgAspArgGlyPro 340
QY 183 CGAAGTCGTGCTCCTCATGCG- - - - -GTGCCA 212
    ||||| ||||| ||||| ||||| |||||
Db 341 GlySerSerArgProAlaAspAlaArgGlyLysArgLysGlnGlyGlyThrProProPro 360
QY 213 CGCCATGGACCTTCTGTCTGTCAGGCCCACTAGTAGGAGGAAGGCGGAGGAG 272
    ||||| ||||| ||||| ||||| |||||
Db 361 SerPro- - - - -ProArgGlyGly- - - - -GlyAlaAlaArgAlaAsnSer 373
QY 273 TGGAGGGGCTCAGCGAAGCTGGGTGCTGTGGGGGTATCCGAGTCCACAGAGCACCTG 332
    ||||| ||||| ||||| ||||| |||||
Db 374 ArgArgProGluGlyAlaAlaProThr- - - - - 382
QY 333 GAACCCCGACAGAAGATTCTGGACTCCCGACGCGGACAGGAGGAGGCGGATGAGCG 392
    ||||| ||||| ||||| ||||| |||||
Db 383 - - - - -SerGlnProGluGlyGluArgLysLysLysLysLysLysLysLysLysLys 395
QY 393 ACACACAAACACAGAACACAC- - - - -AGCCAGTCCCGAGGAGGCCAGT 437
    ||||| ||||| ||||| ||||| |||||
Db 396 LysMetIlyGluThrGluProSerArgGlyAsnLeuIleSerProLysLysLysLysLys 415
QY 438 ANTGGAGAGCCCAAAAGAAAGAACACAGCAGCTGAAAGTCGGGATCCTACACCTGGCGAG 497
    ||||| ||||| ||||| ||||| |||||
Db 416 AsnArgThrProArgSerGluIleProSerArgProSerArgHisSerLysSerGlyGln 435
QY 498 CAGCACAGAAGATCAGGATACAGTCAGATCCCATGTCGCGGACATGGAGGTGATCTG 557
    ||||| ||||| ||||| ||||| |||||
Db 436 SerGluAlaGluAspProAlaAlaAlaGlu- - - - -AlaArgArgArgGluSerAspArg 453
QY 558 CAAGAG- - - - -CTGCATCAGTCNAACACCGGGGATAAATCTGGATTGGGTTC 605
    ||||| ||||| ||||| ||||| |||||
Db 454 ArgGluAlaAlaAspArgLeuArgGluAlaGluAlaAlaGluAlaAlaGluAlaAla 472
QY 606 CGCGCTCAAGGTGAAGATATACCTAAAGAGGAACACTGTAAATGCCAGAGCAGGTGA 665
    ||||| ||||| ||||| ||||| |||||
Db 473 AlaArgGlnAlaGluGlu-IleAlaArgGluGluAlaAlaAlaArgAlaArgGlnAlaGlu 492
QY 666 A 666
Db 492 u 492

RESULT 36
Q95KU4
```

```
ID Q95KU4 PRELIMINARY; PRT: 1422 AA.
AC Q95KU4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TREACLE.
GN TCOF1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Haworth K.E., Islam I., Breen M., Putt W., Binns M., Hopkinson D.,
RA Edwards Y.;
RT *Canine TCOF1: Cloning, Chromosome Assignment and Genetic analysis in
RT dogs with different face and head types.*;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ296287; CAC82741.1; -.
SQ SEQUENCE 1422 AA; 145804 MW; B7381D3EDAF79132 CRC64;

Alignment Scores:
Pred. No.: 0.688 Length: 1422
Score: 98.50 Matches: 55
Percent Similarity: 40.10% Conservative: 22
Best Local Similarity: 28.65% Mismatches: 74
Query Match: 7.14% Indels: 41
DB: 6 Gaps: 9

US-09-658-824-808 (1-781) x Q95KU4 (1-1422)
QY 19 GCGAGTCGGTCCCTGAGGTCGTGGATTCTTCGCTACTGAGACACGCGGGTAGGT 78
    ||||| ||||| ||||| ||||| |||||
Db 245 GlyLysAlaGlyPro- - - - -SerAlaAlaGlnAlaLysThrGluLys 258
QY 79 CCACAGCAGCATCCAACTGGGAGTGTGAAGTGTGAGTGAGAGTGAAGAGAACACGAGGC 138
    ||||| ||||| ||||| ||||| |||||
Db 259 ProLysGluAspSerAspSerSerGluGluAspSerSerGluGluGluProProAla 278
QY 139 TTCGGAGGTTGTGTGGTCACTGAGTCAGTGAGAGGCCCTCGAAGTCGTGCTCCCT 198
    ||||| ||||| ||||| ||||| |||||
Db 279 AlaLysThrProLeu- - - - -GlnValLysProSerGlyLysThrProGlnValLysAlaAla 297
QY 199 CTCATCGGTGGCCACGCCCATGGACCTTCTGCTCGTCAGGCCCATACTAGGAGGAA 258
    ||||| ||||| ||||| ||||| |||||
Db 298 SerAlaSerAla- - - - -LysGluSerProArgLysGlyValProProValProGlyLys 316
QY 259 GGAGGCGCGAGGAGTGAGGGGCTCAGGGGAAGCTGGGTGCTGTGGGGGTATCCGAGT 318
    ||||| ||||| ||||| ||||| |||||
Db 317 ValGlyProAlaAlaGly- - - - -GlnAlaLysLys-GlyAla- - - - -Gln 329
QY 319 CCCAGAGCAGCTGGAAACCCCGACAGAAGATTCTGGAATCTCCAGAC- - - - - 365
    ||||| ||||| ||||| ||||| |||||
Db 329 yGluGluAspProAspSerSerThrGluGluSer-AspSerGluGluGluAlaProThrA 349
QY 366 - - - - -GGGACCAGGAGGAGGCGGATGAGCGGACACACACACACACAGACACACA 417
    ||||| ||||| ||||| ||||| |||||
Db 349 lavalProProThrArg- - - - -SerProValGlnAlaLysProSerG 363
QY 418 GCCAGTCCAGGAG- - - - -CCAGTAATGGAGAGCCCAAGAAAGAG 459
    ||||| ||||| ||||| ||||| |||||
Db 363 yGlnAsnSerGlnValArgThrAlaSerGlyProValLysGlyProGlnLysAlaG 383
QY 460 ACCAGCAGCTGAAGTCCGGATCCCTACCTGGGAGGAGACAGACAGAAGATCAGGATA 519
    ||||| ||||| ||||| ||||| |||||
Db 383 yProAlaAla- - - - -ThrProValGlyLysGlnGluGluAspSerGluS 398
QY 520 CAGCTGAGATCCAGTGGCGGACATGGAAGGT 551
    ||||| ||||| ||||| ||||| |||||
Db 398 erSerSerGluGluSerAspSerGluGly 408

RESULT 37
Q13059
```



RT cation channel that mediates inhibition by calcium-calmodulin.";  
RL J. Biol. Chem. 0:0-0(1998).  
DR EMBL; AF042498; AAC04830.1; -.  
DR InterPro; IPR000636; Cation\_chan\_non\_lig.  
DR InterPro; IPR001622; Channel\_pore\_K.  
DR InterPro; IPR000595; cNMP\_binding.  
DR Pfam; PF00027; cNMP\_binding; 1.  
DR Pfam; PF00520; ion\_trans; 1.  
DR SMART; SM00100; cNMP; 1.  
DR PROSITE; PS00888; CNMP\_BINDING\_1; 1.  
DR PROSITE; PS00889; CNMP\_BINDING\_2; 1.  
DR PROSITE; PS00042; CNMP\_BINDING\_3; 1.  
KW Receptor.  
SQ SEQUENCE 1245 AA; 139159 MW; 40C4860BFCF86126 CRC64;  
  
Alignment Scores:  
Pred. No.: 0.754 Length: 1245  
Score: 98.00 Matches: 33  
Percent Similarity: 43.59% Conservative: 18  
Best Local Similarity: 28.21% Mismatches: 47  
Query Match: 7.10% Indels: 19  
DB: 4 Gaps: 3  
  
US-09-658-824-808 (1-781) x O43636 (1-1245)  
QY 342 CAGAAGATTCTGGACTCCCGGACGAGGAGGCGGATGAGCGACACACA 401  
Db 6 GlnArgValLeuProGluProGlyThrProArgLysThrLysMetGlnGluGlu 25  
QY 402 AACACAGAACACACAGCGATCCCGAGGAGCCAGTAAATGAGAGCCCAAGAGAGAA 461  
Db 26 GluValGluProGluProGluMetGluAlaGluValGluProGluProAsnProGluGlu 45  
QY 462 CCAGCAGCTGAAAGTCGG----- 479  
Db 46 AlaGluThrGluSerGluSerMetProGluGluSerPheLysGluGluValAla 65  
QY 480 -----GATCCTACACCTGGCGGACGACAGAGAGATCAGGATACAGCTGAGATCCCA 533  
Db 66 ValAlaAspProSerPro---GlnGluThrLysGluAlaAlaLeuThrSerThrIleSer 84  
QY 534 GTGCGGACATGAAGGTGATCTGCAAGAGCTGATCAGTCAACACCGGGGATAAATCT 593  
Db 85 LeuArgAlaGlnGlyAlaGluIleSerGluMetAsn-SerProSerHisArgValLeuTh 104  
QY 594 GGATTGGTTCGGCGTCAAGGTGAAGTAATACCTAAAGAGGAACAC 642  
Db 104 rTrpLeuMetLysGlyValGlu---LysValIleProGlnProValHis 119  
  
RESULT 40  
QYUMG2 PRELIMINARY; PRT; 1251 AA.  
AC QYUMG2;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update).  
DE CGMP-GATED CATION CHANNEL BETA SUBUNIT.  
GN CNCG2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE OF 1-291 FROM N.A.  
RC TISSUE=RETINA;  
RX MEDLINE=96070429; PubMed=7590744;  
RA Ardell M.D., Makhija A.K., Oliveira L., Minliou P.,  
RA Viegas-Pequignot E., Pittler S.J.;  
RT "cDNA, gene structure, and chromosomal localization of human GAR1  
RT (CNCG3L), a homolog of the third subunit of bovine photoreceptor cGMP-  
RL gated channel.";  
RL Genomics 28:32-38(1995).  
RN [2]

RP SEQUENCE FROM N.A.  
RC TISSUE=RETINA;  
RX MEDLINE=96338110; PubMed=8766832;  
RA Ardell M.D., Aragon I., Oliveira L., Porche G.E., Burke E.,  
RA Pittler S.J.;  
RT "The beta subunit of human rod photoreceptor cGMP-gated cation channel  
RT is generated from a complex transcription unit.";  
RL FEBS Lett. 389:213-218(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=RETINA;  
RA Pittler S.J.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U58837; AAB63387.1; -.  
DR InterPro; IPR000636; Cation\_chan\_non\_lig.  
DR InterPro; IPR001622; Channel\_pore\_K.  
DR InterPro; IPR000595; cNMP\_binding.  
DR Pfam; PF00027; cNMP\_binding; 1.  
DR Pfam; PF00520; ion\_trans; 1.  
DR SMART; SM00100; cNMP; 1.  
DR PROSITE; PS00888; CNMP\_BINDING\_1; 1.  
DR PROSITE; PS00889; CNMP\_BINDING\_2; 1.  
DR PROSITE; PS00042; CNMP\_BINDING\_3; 1.  
SQ SEQUENCE 1251 AA; 139604 MW; 4701C53DB13C9055 CRC64;  
  
Alignment Scores:  
Pred. No.: 0.754 Length: 1251  
Score: 98.00 Matches: 33  
Percent Similarity: 43.59% Conservative: 18  
Best Local Similarity: 28.21% Mismatches: 47  
Query Match: 7.10% Indels: 19  
DB: 4 Gaps: 3  
  
US-09-658-824-808 (1-781) x Q9UMG2 (1-1251)  
QY 342 CAGAAGATTCTGGACTCCCGGACGAGGAGGCGGATGAGCGACACACA 401  
Db 6 GlnArgValLeuProGluProGlyThrProArgLysThrLysMetGlnGluGlu 25  
QY 402 AACACAGAACACACAGCGATCCCGAGGAGCCAGTAAATGAGAGCCCAAGAGAGAA 461  
Db 26 GluValGluProGluProGluMetGluAlaGluValGluProGluProAsnProGluGlu 45  
QY 462 CCAGCAGCTGAAAGTCGG----- 479  
Db 46 AlaGluThrGluSerGluSerMetProGluGluSerPheLysGluGluValAla 65  
QY 480 -----GATCCTACACCTGGCGGACGACAGAGAGATCAGGATACAGCTGAGATCCCA 533  
Db 66 ValAlaAspProSerPro---GlnGluThrLysGluAlaAlaLeuThrSerThrIleSer 84  
QY 534 GTGCGGACATGAAGGTGATCTGCAAGAGCTGATCAGTCAACACCGGGGATAAATCT 593  
Db 85 LeuArgAlaGlnGlyAlaGluIleSerGluMetAsn-SerProSerHisArgValLeuTh 104  
QY 594 GGATTGGTTCGGCGTCAAGGTGAAGTAATACCTAAAGAGGAACAC 642  
Db 104 rTrpLeuMetLysGlyValGlu---LysValIleProGlnProValHis 119  
  
Search completed: October 18, 2002, 09:42:57  
Job time : 73 secs